PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : G01N 33/569, C07K 13/00, C12N 15/30 — A1

(11) International Publication Number: WO 95/00849

(43) International Publication Date: 5 January 1995 (05.01.95)

(21) International Application Number: PCT/US94/06890

(22) International Filing Date: 17 June 1994 (17.06.94)

(30) Priority Data:
08/078,476 17 June 1993 (17.06.93) US
08/130,735 1 October 1993 (01.10.93) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/130,735 (CIP)
Filed on 1 October 1993 (01.10.93)

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF VIRGINIA PATENT FOUNDATION [US/US]; 1224 West Main Street, Charlottesville, VA 22903 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MANN, Barbara, J. [US/US]; 58 Woodlake Drive, Charlottesville, VA 22901 (US). PETRI, William, A. [US/US]; 2584 Holkham Drive, Charlottesville, VA 22901 (US).

(74) Agents: MURASHIGE, Kate, H. et al.; Morrison & Foerster, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006 (US).

(81) Designated States: AU, BB, BG, BR, BY, CA, CN, CZ, FI, HU, JP, KP, KR, KZ, LK, LV, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: RECOMBINANT 170 KD SUBUNIT LECTIN OF ENTAMOEBA HISTOLYTICA AND METHODS OF USE

(57) Abstract

Recombinantly produced 170 kDa heavy subunit of *Entamoeba histolytica* Gal/GalNAc adherence lectin or an epitope-bearing portion thereof may be used as antigen in serological analysis of *E. histolytica* infection or as a vaccine for immunizing against infection. The recombinant production is achieved in a procaryotic system to provide non-glycosylated antigens or immunogens which are immunologically reactive. Also disclosed are three genes which encode epitope-bearing portions of the lectin.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

RECOMBINANT 170 KD SUBUNIT LECTIN OF ENTAMOEBA HISTOLYTICA AND METHODS OF USE

This invention was made, in part, with support supplied by the U.S. Government under Contracts AI 18841 and AI 26649 awarded by the National Institutes of Health. The U.S. Government has certain rights in this invention.

Field of the Invention

The invention concerns the use of epitope-bearing regions of the 170 kD subunit of *Entamoeba histolytica* Gal/GalNAc adherence lectin which are produced recombinantly in procaryotic systems in diagnosis and as vaccines. Thus, the invention relates to the determination of the presence, absence or amount of antibodies raised by a subject in response to infection by *E. histolytica* using these peptides and to vaccines incorporating them. This invention also particularly relates to reagents specific for a novel variant of the 170 kD subunit of *E. histolytica* Gal/GalNAc adherence lectin and to the gene (*hgl3*) which encodes this novel subunit form, which represents the third member of the multigene family encoding this 170 kD subunit.

Background Art

Entamoeba histolytica infection is extremely common and affects an estimated 480 million individuals annually. However, only about 10% of these persons develop symptoms such as colitis or liver abscess. The low incidence of symptom occurrence is putatively due to the existence of both pathogenic and nonpathogenic forms of the amoeba. As of 1988, it had been established that

the subjects who eventually exhibit symptoms harbor pathogenic "zymodemes" which have been classified as such on the basis of their distinctive hexokinase and phosphoglucomutase isoenzymes. The pathogenic forms are not conveniently distinguishable from the nonpathogenic counterparts using morphogenic criteria, but there is an almost perfect correlation between infection with a pathogenic zymodeme and development of symptoms and between infection with a nonpathogenic zymodeme and failure to develop these symptoms.

It is known that *E. histolytica* infection is mediated at least in part by the "Gal/GalNAc" adherence lectin which was isolated from a pathogenic strain and purified 500 fold by Petri, W.A., et al., J Biol Chem (1989) 264:3007-3012. The purified "Gal/GalNAc" lectin was shown to have a nonreduced molecular weight of 260 kD on SDS-PAGE; after reduction with beta-mercaptoethanol, the lectin separated into two subunits of 170 and 35 kD MW. Further studies showed that antibodies directed to the 170 kD subunit were capable of blocking surface adhesion to test cells (Petri, et al. J Biol Chem (1989) supra). Therefore, the 170 kD subunit is believed to be of primary importance in meditating adhesion.

In addition, the 170 kD subunit is described as constituting an effective vaccine to prevent *E. histolytica* infection in U.S. Patent 5,004,608 issued 2 April 1991.

Studies of serological cross-reactivity among patients having symptomology characteristic of *E. histolytica* pathogenic infection, including liver abscess and colitis, showed that the adherence lectin was recognized by all sera tested (Petri, Jr., W.A., et al., Am J Med Sci (1989) 296:163-165). The lectin heavy subunit is almost universally recognized by immune sera and T-cells from patients with invasive amebiasis (Petri,

et al., Infect Immun (1987) 55:2327-2331; Schain, et al., Infect Immun (1992) 60:2143-2146).

DNA encoding both the heavy (170 kD) and light (35 kD) subunits have been cloned. The heavy and light subunits are encoded by distinct mRNAs (Mann, B., et al., Proc Natl Acad Sci USA (1991) 88:3248-3252) and these subunits have different amino acid compositions and amino terminal sequences. The sequence of the cDNA encoding the 170 kD subunit suggests it to be an integral membrane protein with a large cysteine-rich extracellular domain and a short cytoplasmic tail (Mann, B., et al., Proc Natl Acad Sci USA (1991) supra; Tannich, et al., Proc Natl Acad Sci USA (1991) 88:1849-1853). The derived amino acid sequence of the 170 kD lectin shows that the extracellular domain can be divided into three regions on the basis of amino acid composition. The amino terminal amino acids 1-187 are relatively rich in cysteine (3.2%) and tryptophan (2.1%). Amino acid sequence at positions 188-378 does not contain cysteine, and the amino acid sequence at positions 379-1209 contains 10.8% cysteine residues. The obtention of clones encoding the heavy chain subunit is further described in U.S. Patent 5,260,429 issued 9 November 1993, the disclosure of which is incorporated herein by reference. In that patent, diagnostic methods for the presence of *E. histolytica* based on the polymerase chain reaction and the use of DNA probes is described.

The heavy subunit is considered to be encoded by a multigene family (Mann, B., et al., Parasit Today (1991) 1:173-176). Two different heavy subunit genes, *hgl1* and *hgl2*, have been sequenced by separate laboratories. While *hgl2* was isolated from an HM-1:IMSS cDNA library in its entirety (Tannich, E. *et al.* Proc Natl Acad Sci USA (1991) 88:1849-1853), *hgl1* was isolated in part from an H-302:NIH cDNA library and in part by PCR amplification of the gene from the HM-1:IMSS genome

(Mann, B.J. *et al.* Proc Natl Acad Sci USA (1991) 88:3248-3252). As the amino acid sequence of these two genes is 87.6% identical (Mann, B.J. *et al.* Parasit Today (1991) 7:173-176), the differences could be explained by strain variation alone. The presence of multiple bands hybridizing to an *hgl* probe on Southern blots, however, in consistent with the existence of a 170 kDa subunit gene family (Tannich, E. *et al.* Proc Natl Acad Sci USA (1991) 88:1849-1853).

Monoclonal antibodies specifically immunoreactive with various epitope-bearing regions of the 170 kD heavy chain subunit have also been disclosed in U.S. Patent 5,272,058 issued 21 December 1993, the disclosure of which is incorporated herein by reference in its entirety. This application also describes use of these antibodies to detect the 170 kD heavy chain and the use of the 170 kD subunit to detect antibodies in serum or other biological samples. The experimental work described utilizes the native protein. Further characterization of these antibodies is described in a publication by Mann, B.J., et al., Infect Immun (1993) 61:1772-1778 also incorporated herein by reference.

Various immunoassay techniques have been used to diagnose *E. histolytica* infection. ELISA techniques have been used to detect the presence or absence of *E. histolytica* antigens both in stool specimens and in sera, though these tests do not seem to distinguish between the pathogenic and nonpathogenic strains. In a seminal article, Root, et al., Arch Invest Med (Mex) (1978) 9: Supplement 1:203, described the use of ELISA techniques for the detection of amoebic antigen in stool specimens using rabbit polyclonal antiserum, and various forms of this procedure have been used, some in conjunction with microscopic studies. Palacios et al., Arch Invest Med (Mex) (1978) 9: Supplement 1:203; Randall et al., Trans Roy Soc Trop Med Hyg (1984) 78:593; Grundy,

Trans Roy Soc Trop Med Hyg (1982) 76:396; Ungar, Am J Trop Med Hyg (1985) 34:465. These studies on stool specimens and on other biological fluids are summarized in Amebiasis: Human Infection by Entamoeba Histolytica, J. Ravdin, ed. (1988) Wiley Medical Publishing, pp. 646-648.

Conversely, amebic serology is also a critical component in the diagnosis of invasive amebiasis. One approach utilizes conventional serologic tests, such as the indirect hemagglutinin test. These tests are very sensitive but seropositivity is persistent for years (Krupp, I.M., Am J Trop Med Hyg (1970) 19:57-62; Lobel, H.O. et al., Ann Rev Microbiol (1978) 32:379-347). Thus, healthy subjects may give positive responses to the assay, creating an undesirable high background. Similar problems with false positives are found in using immunoassay tests involving a monoclonal antibody and purified native 170 kD protein (Ravdin, J.I., et al., J Infect Dis (1990) 162:768-772.)

Recombinant *E. histolytica* proteins other than the 170 kD subunit have been used as the basis for serological tests. Western blotting using a recombinant form of the "52 kD serine-rich protein" was highly specific for invasive disease and had a higher predictive value (92 vs. 65%) than an agar gel diffusion test for diagnosis of acute amebiasis (Stanley, Jr., S.L., et al., Proc Natl Acad Sci U.S.A. (1990) 87:4976-4980; Stanley, Jr., S.L., et al., JAMA (1991) 266:1984-1986). However, the overall sensitivity was lower than for the conventional agar gel test (82% vs. 90-100%).

Thus, there remains a need for serological tests which will provide optimum sensitivity while minimizing the number of false positives retained. The present invention provides such a test by utilizing, as antigen, epitope-bearing portions of the 170 kD subunit

of the adherence lectin produced recombinantly in procaryotic systems.

It is particularly advantageous to use recombinantly produced, nonglycosylated peptides or proteins in this assay since these peptides are easily and efficiently obtained and are easily standardized. Furthermore, since selected portions of the lectin heavy chain subunit can be produced, epitopes characteristic of the pathogenic or nonpathogenic forms of *E. histolytica* can be produced and used to distinguish these forms in the assays. Subsequent to the invention herein, a report of immunoreactivity of recombinant 170 kd lectin with immune sera was published by Zhang, Y, et al. J. Clin Micro-immunol (1992) 2788-2792. Applicants incorporate by reference their own publication: Mann, B.J et al. Infect and Immun (1993) 61: 1772-1778.

Similarly, although it is known that the 170 kD subunit may be used as a vaccine as described in the above-referenced U.S. Patent 5,004,608, recombinantly produced forms of the 170 kD subunit, specifically those obtained from procaryotic cells that lack glycosylation may offer advantages in reproducibility of product and in ease of preparation of subunit vaccines. The present invention is directed to this desirable result.

Disclosure of the Invention

The invention provides diagnostic tests which permit the assessment of patients for invasive *E. histolytica* infection and vaccines for prevention of infection. The invention also provides a novel third variant of the 170 kD subunit of the Gal/GalNAc adherence lectin and a gene (*hgl3*) which encodes this novel protein. Accordingly, the diagnostic tests of the invention are based on the genetic sequences of all three variants of the 170 kD subunit of the Gal/GalNAc

adherence lectin which are encoded by three different genes in a multigene family.

Pathogenic and nonpathogenic strains can be distinguished by use of the invention diagnostic method, if desired. The tests use, as antigen, an epitope-bearing portion of the 170 kD subunit of the Gal/GalNAc adherence lectin recombinantly produced in procaryotic systems. Despite the absence of glycosylation from such portions and despite the lack of post-translational modifications characteristic of the native protein or peptide, the recombinantly produced proteins are effective antigens in these assays.

Thus, in one aspect, the invention is directed to a method to detect the presence or absence of antibodies immunoreactive with pathogenic and/or nonpathogenic *E. histolytica* in a biological sample which method comprises contacting the fluid with an epitope-bearing portion of the 170 kD heavy chain of the Gal/GalNAc adherence lectin wherein the lectin is nonglycosylated and in a form obtainable from procaryotic cells. If distinction between antibodies to the pathogenic and nonpathogenic forms is desired, the portion may be chosen so as to be characteristic of the pathogenic or nonpathogenic form. Alternatively, the assay may be conducted as a competition assay using MAbs with such characteristics. The contacting is conducted under conditions where the epitope-bearing portion forms complexes with any antibodies present in the biological fluid which are immunoreactive with an epitope on the portion. The presence, absence or amount of such complexes is then assessed, either directly or in a competition format, as a measure of the antibody contained in the biological sample. The invention is also directed to materials and kits suitable for performing the methods of the invention.

In a second aspect, the invention is directed to methods to prevent *E. histolytica* infection using vaccines containing, as active ingredient, epitope-bearing portions of the 170 kD subunit produced recombinantly in procaryotic systems, as described above. The invention is also directed to vaccines containing this active ingredient.

In other aspects, the invention is directed to epitope-bearing portions of the 170 kD subunit produced recombinantly in procaryotic systems and thus in a form characteristic of such production. One characteristic is lack of glycosylation; in addition, secondary structure of proteins produced by procaryotic hosts differs from that of proteins produced by the natural source.

In yet another aspect, the invention is directed to a DNA in purified and isolated form which consists essentially of a DNA encoding the 170 kd heavy chain subunit of pathogenic *E. histolytica* Gal/GalNAc adherence lectin, which subunit is encoded by the *hgl3* gene for which the nucleotide sequence and deduced amino acid sequence are shown in Figure 4. In further aspects, the invention is directed to both nucleic acid and immunological reagents which are enabled by the discovery of the *hgl3* gene, reagents which are specific for each of the *hgl1*, *hgl2* or *hgl3* genes, as well as reagents which detect common regions of all three *hgl* genes or their nucleic acid or protein products. For example, oligonucleotide probes specific for any one of these three genes or for a sequence common to all three genes may be identified by one of ordinary skill in the art, using conventional nucleic acid probe design principles, by comparisons of the three DNA sequences for these genes. See Example 6.

In still further aspects, the invention is directed to a method to detect the presence, absence, or amount of a pathogenic or nonpathogenic form of Entamoeba

histolytica, where *E. histolytica* has both pathogenic and nonpathogenic forms, in a biological sample, which method comprises contacting the sample with a monoclonal antibody immunospecific for an epitope of the 170 kd subunit of Gal/GalNAc lectin unique to the pathogenic or to the nonpathogenic form, or shared by the pathogenic and nonpathogenic forms of *E. histolytica*, to form an immunocomplex when the pathogenic and/or nonpathogenic form is present, and detecting the presence, absence or amount of the immunocomplex. In this method, the epitope is selected to be specific for one of 170 kD subunits encoded by the *hgl*1, *hgl*2 or *hgl*3 genes, or for a common region of the subunits from all three *hgl* genes.

In another aspect, the invention is directed to a method to determine the presence, absence or amount of antibodies specifically immunoreactive with the Gal/GalNAc lectin derived from *E. histolytica*, which method comprises contacting a biological sample with the Gal/GalNAc lectin or the 170 kd subunit thereof in purified and isolated form, under conditions wherein antibodies immunospecific for said lectin or subunit will forma complex, and detecting the presence, absence or amount of the complex, wherein the purified and isolated Gal/GalNAc lectin or subunit is derived from either a pathogenic or nonpathogenic form of *E. histolytica*, and is a 170 kD subunit encoded by one of the *hgl*1, *hgl*2 or *hgl*3 genes. Detailed descriptions of these and related methods for detecting pathogenic or nonpathogenic forms of *E. histolytica* and antibodies specifically immunoreactive with the Gal/GalNAc lectin derived from *E. histolytica*, as well as reagent kits suitable for the conduct of such methods, are disclosed in U.S. Patent 5,272,058, the entire disclosure of which is incorporated herein by reference.

Brief Description of the Drawings

Figure 1A shows the DNA and amino acid sequence deduced from the nucleotide sequence corresponding to the 170 kD heavy chain of the adherence lectin from pathogenic strain HM1:IMSS, designated *hgl1*.

Figure 1B shows the deduced amino acid sequence of *hgl1* with the amino-terminal amino acid of the mature protein designated as amino acid number 1.

Figure 2A is a diagram of the construction of expression vectors for recombinant production of specified portions of the 170 kD subunit; Figure 2B shows the pattern of deletion mutants.

Figure 3 is a diagram of the location of human B cell epitopes and pathogenic-specific epitopes on the 170 kD heavy chain.

Figure 4A shows the DNA and amino acid sequence deduced from the nucleotide sequence corresponding to the 170 kD heavy chain of the adherence lectin from pathogenic strain HM1:IMSS, designated *hgl3*.

Figure 4B shows the deduced amino acid sequence of *hgl3* with the amino-terminal amino acid of the mature protein designated as amino acid number 1. The putative signal sequence and transmembrane domains are overlined and underlined respectively. Conserved cysteine residues (●) and potential sites of glycosylation (*) are indicated.

Figure 5 shows in schematic form a comparison of amino acid sequences of three heavy subunit genes. The top diagram represents a schematic representation of a heavy subunit gene. Starting at the amino terminus, regions include the cysteine/tryptophan (C-W) rich domain, the cysteine-free (C-free) domain, the cysteine-rich (C-rich) domain, and the putative transmembrane (TM) sequence and cytosolic domains (Mann, B.J. *et al.* Parasit Today (1991) 7:173-176). Amino acid sequence comparisons of *hgl1*, *hgl2* and *hgl3* are shown. Upright lines indicate

nonconservative amino acid substitutions in the amino acid sequence of the second gene as compared to the first gene listed to the right. Downward arrowheads indicate a deletion while upright arrowheads indicate an insertion. The number of residues inserted or deleted are listed below the arrowheads and the total percent amino acid sequence identity is listed at right.

Modes of Carrying Out the Invention

The invention provides methods and materials which are useful in assays to detect antibodies directed to pathogenic and/or nonpathogenic forms of *E. histolytica* and in vaccines. The diagnostic assays can be conducted on biological samples derived from subjects at risk for infection or suspected of being infected. The assays can be designed to distinguish pathogenic from nonpathogenic forms of the amoeba if desired. The vaccines are administered to subjects at risk for amebic infections.

The assays of the invention rely on the ability of an epitope-bearing portion of the 170 kD subunit produced recombinantly in procaryotic cultures to immunoreact with antibodies contained in biological samples obtained from individuals who have been infected with *E. histolytica*. Even though the relevant peptide or protein is produced in a procaryotic system, and is thus not glycosylated or processed after translation in a manner corresponding to the native protein, the epitope-bearing portions thus prepared are useful antigens in immunoassays performed on samples prepared from biological fluids, cells, tissues or organs, or their diluted or fractionated forms. Similarly, these peptides are also immunogenic.

The use of recombinant forms of the antigen or offers advantages of cost-effective, reliable production of pure antigen, thus assuring the uniformity of the

assay materials. Recombinant production in bacteria is a particularly efficient and useful method. It is surprising that such procaryotic systems can produce successful antigens and immunogens, since the peptides produced are not processed in a manner analogous to the reactive native forms.

Furthermore, recombinant production facilitates the preparation of specific epitopes, thus providing a means for detecting antibodies specifically immunoreactive with pathogenic or nonpathogenic forms of the amoeba, as well as offering the opportunity to provide subunit vaccines.

Thus, the invention is directed to methods to detect antibodies in biological samples and to immunize subjects at risk using these recombinantly produced epitope-bearing portions as antigens or immunogens as well as to the recombinantly produced peptides themselves and to materials useful in performing the assays and in administering the vaccines.

Definitions

The diagnostic assays may be designed to distinguish antibodies raised against nonpathogenic or pathogenic forms of the amoeba. "Pathogenic forms" of *E. histolytica* refers to those forms which are invasive and which result in symptomology to infected subjects. "Nonpathogenic forms" refers to those forms which may be harbored asymptomatically by carriers.

The assays and vaccines of the invention utilize an epitope-bearing portion of the 170 kD subunit of the Gal/GalNAc lectin. "Gal/GalNAc lectin" refers to glycoprotein found on the surface of *E. histolytica* which mediates the adherence of the amoeba to target cells, and which mediation is inhibited by galactose or N-acetylgalactosamine. The Gal/GalNAc lectin refers specifically to the lectin reported and isolated by

Petri, et al. (supra) from the pathogenic strain HMI-IMSS, and to the corresponding lectin found in other strains of *E. histolytica*. The "170 kD subunit" refers to the large subunit, upon reduction of the Gal/GalNAc lectin, such as that obtained by Petri, et al. and shown in Figure 1 as well as to its corresponding counterparts in other strains.

Diagnostic Assays

With respect to the diagnostic assays of the invention, the complete 170 kD antigen or an epitope-bearing portion thereof can be used in the assays. Such epitope-bearing portions can be selected as characteristic of pathogens or nonpathogens or common to both.

As shown hereinbelow, the portion of the 170 kD protein which contains epitopes for all monoclonal antibodies prepared against the lectin is found at amino acid positions 596-1138. There appears to be an epitope characteristic of pathogens between each of amino acid positions 596-818, 1082-1138, and 1033-1082. Positions 895-998 contain epitopes which are shared by pathogens and nonpathogens as well as epitopes characteristic of pathogenic strains. Thus, to utilize fragments of the recombinantly produced protein for detection of antibodies, a peptide representing positions 596-818, 1033-1082 or 1082-1138 may be used to detect antibodies raised against pathogens by hosts in general; however, the epitope at positions 596-816 is not recognized by human antisera. Mixtures of these peptides could also be used. Alternatively, longer forms of the antigen can be used by selecting the appropriate positions depending on whether pathogenic and nonpathogenic amoebae are to be distinguished.

As shown in Example 4, below, epitope-bearing portions relevant for human testing include portions 2-

482, 1082-1138, 1032-1082 and 894-998. Only the portions represented by 1082-1138 and 1032-1082 appears specific for antibodies against pathogenic ameba. These epitope-bearing portions may be used as single peptides, as uniquely lectin-derived portions of chimeric proteins, as mixtures of peptides or of such proteins, or as portions of a single, multiple-epitope-bearing protein. Procedures for preparing recombinant peptide proteins containing only a single epitope-bearing portion identified above, or multiples of such portions (including tandem repeats) are well understood in the art.

The assays are designed to detect antibodies in biological samples which are "immunospecific" or "immunoreactive" with respect to the epitope-bearing portion -- i.e. with respect to at least one epitope contained in this portion. As used herein, "immunospecific" or "immunoreactive" with respect to a specified target means that the antibody thus described binds that target with significantly higher affinity than that with which it binds to alternate haptens. The degree of specificity required may vary with circumstances, but typically an antibody immunospecific for a designated target will bind to that target with an affinity which is at least one or two, or preferably several orders or magnitude greater than with which it binds alternate haptens.

The assays can be performed in a wide variety of protocols depending on the nature of the sample, the circumstances of performing the assays, and the particular design chosen by the clinician. The biological sample is prepared in a manner standard for the conduct of immunoassays; such preparation may involve dilution if the sample is a biological fluid, fractionation if the sample is derived from a tissue or organ; or other standard preparation procedures which are

known in the art. Thus, "biological sample" refers to the sample actually used in the assay which is derived from a fluid, cell, tissue or organ of a subject and prepared for use in the assay using the standard techniques. Normally, plasma or serum is the source of biological sample in these assays.

The assays may be conducted in a competition format employing a specific binding partner for the epitope-bearing portion. As used herein, "specific binding partner" refers to a substance which is capable of specific binding to a targeted substance, such as the epitope-bearing portion of the 170 kD subunit. In general, such a specific binding partner will be an antibody, but any alterative substance capable of such specific binding, such as a receptor, enzyme or arbitrarily designed chemical compound might also be used. In such contexts, "antibody" refers not only to immunoglobulin per se, but also to fragments of immunoglobulin which retain the immunospecificity of the complete molecule. Examples of such fragments are well known in the art, and include, for example, Fab, Fab', and $F(ab')_2$ fragments. The term "antibody" also includes not only native forms of immunoglobulin, but forms of the immunoglobulin which have been modified, as techniques become available in the art, to confer desired properties without altering the immunospecificity. For example, the formation of chimeric antibodies derived from two species is becoming more practical. In short, "antibodies" refers to any component of or derived form of an immunoglobulin which retains the immunospecificity of the immunoglobulin per se.

A particularly useful form of specific binding reagents useful in the assay methods of the invention is as monoclonal antibodies. Three categories of monoclonal antibodies have been prepared to the 170 kD subunit. One category of antibody is immunospecific for epitopes

"unique" to pathogenic forms. These antibodies are capable, therefore, of immunoreaction to a significant extent only with the pathogenic forms of the amoeba or to the 170 kD subunit of lectin isolated from pathogenic forms. A second set of monoclonal antibodies is immunoreactive with epitopes which are "unique" to nonpathogenic forms. Thus, these antibodies are immunoreactive to a substantial degree only with the nonpathogenic amoeba or their lectins and not to the pathogenic forms. A third category of monoclonal antibodies is immunoreactive with epitopes common to pathogenic and nonpathogenic forms and these antibodies are capable of immunoreaction with the subunit or with the amoeba regardless of pathogenicity.

With respect to the monoclonal antibodies described herein, those immunoreactive with epitopes 1 and 2 of the 170 kD subunit isolated from the pathogenic-strain exemplified are capable of reacting, also, with the corresponding epitopes on nonpathogens. On the other hand, those immunoreactive with epitopes 3-6 are capable of immunoreaction only with the 170 kD subunit of pathogenic strains. By applying the techniques for isolation of the pathogenic 170 kD subunit to amoeba which are nonpathogenic, a 170 kD subunit can be obtained for immunization protocols which permit the analogous preparation of MAbs immunoreactive with counterpart epitopes 3-6 in the nonpathogenic forms.

Of course, with respect to antibodies found in the biological sample, in general, these will be found in the form of immunoglobulins. However, pretreatment of the sample with an enzyme, for example, to remove the F_c portions of the antibodies contained therein, does not debilitate the sample with respect to its ability to respond to the assay.

Assay Procedure

For the conduct of the assays of the invention, in general, the biological sample is contacted with the epitope-bearing portion used as an antigen in the immunoassay. The presence, absence or amount of the resulting complex formed between any antibody present in the sample and the epitope-bearing portion is measured directly or competitively.

As is well understood in the art, once the biological sample is prepared, there is a multiplicity of alternative protocols for conduct of the actual assay. In one rather straightforward protocol, the epitope-bearing portion provided as antigen may be coupled to a solid support, either by adsorption or by covalent linkage, and treated with the biological sample. The ability of any antibodies in the sample to bind to coupled antigen is then determined.

This ability may be determined in a "direct" form of the assay in which the level of complex formation by the antibody is measured directly. In one particularly convenient format of this approach, the antigen may be supplied as a band on a polyvinylidene difluoride (PVDF) and contacted with the biological sample; any resulting complexes formed with antibody on the PVDF membrane are then detected as described above for Western blot procedure. This protocol is substantially a Western Blot procedure. Alternatively, microtiter plates or other suitable solid supports may be used. The binding of antibody to the antigen coupled to support can then be detected as described above for Western blot procedure using conventional techniques generally involving secondary labeling using, for example, antibodies to the species from which the biological sample is derived. Such labels may include radioisotopes, fluorescent tags, enzyme labels and the like, as is conventionally understood.

The assay may also be formatted as a competition assay wherein the antigen coupled to solid support is treated not only with the biological sample but also with competing specific binding partner immunospecific for at least one epitope contained in the antigen. The competing binding partner is preferably an antibody. The competing antibody may be polyclonal or monoclonal and may itself be labeled or may be capable of being labeled in a secondary reaction. In a typical conduct of such a competitive test, a competitive specific binding partner for the antigen is generally supplied in labeled form and the success of the competition from the biological sample is measured as a reduction in the amount of label bound in the resulting complex or increased levels of label remaining in the supernatant. If monoclonal antibodies are used, the assay can readily be made specific for pathogenic or nonpathogenic reacting antibodies, if desired, by choosing antibodies of the appropriate specificity. Thus, if the assay is to be made specific for antibodies raised against pathogenic forms of *E. histolytica*, the competition will be provided by a monoclonal antibody specific for an epitope characteristic of pathogenic strains.

Another manner in which the assay may be made specific for pathogenic or nonpathogenic forms is in the choice of the epitope-bearing portion. If antibodies specific to the pathogens are to be detected, an epitope-bearing portion is chosen which bears only epitopes characteristic of pathogenic strains. Conversely, antibodies immunospecific for nonpathogens can be conducted by utilizing as antigen only portions of the subunit which contain epitopes characteristic of nonpathogens. Where characterization as pathogen or nonpathogen-specific antibodies is unnecessary, antigen

containing both such epitopes or epitopes shared by both forms may be used.

Additional ways to distinguish between antibodies immunospecific for pathogens and for nonpathogens employ competition assays with monoclonal antibodies of such specificities, as described above.

Alternatively, the biological sample can be coupled to solid support and the desired epitope-bearing portion added under conditions where a complex can be formed to the epitope-bearing portion, which is then used to treat the support. Subsequent treatment of the support with antibodies known to immunoreact with the antigen can then be used to detect whether antigen has been bound.

Thus, the biological sample to be tested is contacted with the epitope-bearing portion, which is derived either from a pathogenic or nonpathogenic from one both of *E. histolytica* so that a complex is formed. The complex is then detected by suitable labeling, either by supplying the antigen in labeled from, or by a secondary labeling process which forms a ternary complex. The reaction is preferably conducted using a solid phase to detect the formation of the complex attached to solid support, or the complex can be precipitated using conventional precipitating agents such as polyethylene glycol.

In a more complex form of the assay, competitive assays, can be used wherein the biological sample, preferably serum or plasma, provides the cold antibody to compete with a specific binding partner, such as a labeled monoclonal antibody preparation known to bind specifically to an epitope unique to the Gal/GalNAc lectin or its 170 kD subunit of a pathogenic or nonpathogenic from. In this embodiment, the binding to labeled specific monoclonal antibody is conducted in the presence and absence of biological sample, and the

diminution of labeling of the resulting complex in the presence of sample is used as an index to determine the level of competing antibody.

Kits suitable for the conduct of these methods include the appropriate labeled antigen or antibody reagents and instructions for conducting the test. The kit may include the antigen coupled to solid support as well as additional reagents.

Methods of Protection and Vaccines

The recombinant 170 kD subunit or an epitope-bearing portion thereof may be used as active ingredient. Preferred regions include positions 482-1138, 596-1138, 885-998, 1033-1082 and 1082-1138.

The 170 kD subunit or its epitope-bearing regions may also be produced recombinantly in procaryotic cells for the formulation of vaccines. The recombinantly produced 170 kD protein or an epitope-bearing region thereof can be used as an active ingredient in vaccines for prevention of *E. histolytica* infection in subjects who are risk for such condition. Sufficiently large portions of the 170 kD protein can be used per se; if only small regions of the molecules for example containing 20 amino acids or less or to be used, it may advantageous to couple the peptide to a neutral carrier to enhance its immunogenicity. Such coupling techniques are well known in the art, and include standard chemical coupling techniques optionally effected through linker moieties such as those available from Pierce Chemical Company, Rockford, Illinois. Suitable carriers may include, for example, keyhole limpid hemocyanin (KLH) *E. coli* pilin protein k99, BSA, or the VP6 protein of rotavirus. Another approach employs production of fusion proteins which include the epitope-bearing regions fused to additional amino acid sequence. In addition, because of the ease with which recombinant materials can be

manipulated, the epitope-bearing region may be included in multiple copies in a single molecule, or several epitope-bearing regions can be "mixed and matched" in a single molecule.

The active ingredient, or mixture of active ingredients, in the vaccine is formulated using standard formulation for administration of proteins or peptides and the compositions may include an immunostimulant or adjuvant such as complete Freund's adjuvant, aluminum hydroxide, liposomes, ISCOMs, and the like. General methods to prepare vaccines are described in Remingtons's Pharmaceutical Science; Mack Publishing Company Easton, PA (latest edition). The compositions contain an effective amount of the active ingredient peptide or peptides together with a suitable amount of carrier vehicle, including, if desired, preservatives, buffers, and the like. Other descriptions of vaccine formulations are found in "New Trends and Developments in Vaccines", Voller, A., *et al.*, University Park Press, Baltimore, Maryland (1978).

The vaccines are administered as is generally understood in the art. Ordinarily, administration is systemic through injection; however, other effective means of administration are included. With suitable formulation, for example, peptide vaccines may be administered across the mucus membrane using penetrants such as bile salts or fusidic acids in combination, usually, with a surfactant. Transcutaneous means for administering peptides are also known. Oral formulations can also be used. Dosage levels depend on the mode of administration, the nature of the subject, and the nature of carrier/adjuvant formulation. Typical amounts of protein are in the range of .01 μg-1 mg/kg. However, this is an arbitrary range which is highly dependent on the factors cited above. In general, multiple

administrations in standard immunization protocols are preferred; such protocols are standard in the art.

A preferred epitope-bearing region of the 170 kD subunit is that represented by amino acids 482-1138 which includes the cysteine-rich domain. This region is encoded by nucleotides 1492-3460 shown in Figure 1 herein. Preferred regions include those bearing epitopes which are specific for antibodies against pathogenic amoeba -- i.e., regions 1082-1138 and 1032-1082. However, the epitope-bearing region at positions 894-998 may also be used. For regions of this length, production of peptides with multiple copies of the epitope-bearing regions is particularly advantageous.

Production of Recombinant Epitope-bearing Portions

The epitope-bearing portions of the 170 kD subunit can be conveniently prepared in a variety of procaryotic systems using control sequences and hosts ordinarily available in the art. The portions may be provided as fusion proteins or as mature proteins and may be produced intracellularly or secreted. Techniques for constructing expression systems to effect all of these outcomes is well understood in the art. If the epitope-bearing portion is secreted, the medium can be used directly in the assay to provide the antigen, or the antigen can be recovered from the medium and further purified if desired. If the protein is produced intracellularly, lysates of cultured cells may be used directly or the protein may be recovered and further purified. In the Examples below, the epitope-bearing portion is provided as a fusion protein using the commercially available expression vector pGEX. Alternative constructions and alternative hosts can also be used as is understood in the art.

Reagents and assays for a novel 170 kD lectin subunit

To determine the existence and complexity of the 170 kDa subunit gene family, *hgl*, an amebic genomic library in lambda phage was hybridized with DNA fragments from the 5′ or 3′ ends of *hgl*1. Termini from three distinct heavy subunit genes were identified including *hgl*1, *hgl*2, and a third, unreported gene designated *hgl*3. The open reading frame of *hgl*3 was sequenced in its entirety (Figure 4A). Nonstringent hybridization of a genomic Southern blot with heavy subunit specific DNA labeled only those bands predicted by *hgl*1-3. The amino acid sequence of *hgl*3 (Figure 4B) was 95.2% identical to *hgl*1 and 89.4% identical to *hgl*2. All 97 cysteine residues present in the heavy subunit were conserved in *hgl*1-3. Analysis of amebic RNA showed that all three heavy subunit genes were expressed in the amebae and that *hgl* message became less abundant as the amebae entered a stationary growth phase.

Accordingly, the present invention provides both nucleic acid and immunological reagents specific for 170 kDa subunits encoded by each of the *hgl*1, *hgl*2 or *hgl*3 genes, as well as reagents which detect common regions of all three *hgl* genes and their nucleic acid or protein products. For example, oligonucleotide probes specific for any one of these three genes may be identified by one of ordinary skill in the art, using conventional nucleic acid probe design principles, by comparisons of the three DNA sequences for these genes, which sequences are disclosed in Figure 1A and Figure 4A for *hgl*1 and *hgl*3, respectively, and for *hgl*2, in Tannich, E. *et al*. Proc Natl Acad Sci USA (1991) 88:1849-1853, the entire disclosure of which is hereby incorporated herein by reference. Example 6 illustrates the use of oligonucleotide probes specific for each of the three *hgl* genes, for determining the level of expression of RNA from each gene using Northern blot analyses. Other methods of using *hgl*-specific nucleic

acids for diagnostic purposes, for pathogenic and/or nonpathogenic forms of *E. histolytica*, are described in U.S. Patent 5,260,429, the entire disclosure of which is incorporated herein by reference.

The following Examples are intended to illustrate but not to limit the invention.

Example 1

Construction of Expression Vectors

The 170 kD subunit of the galactose lectin is encoded by at least two genes. The DNA used for all of the constructions described herein encodes the 170 kD lectin designated *hgl1* (Fig. 1A). The nucleotide position designations refer to the numbering in Figure 1A.

The DNA sequence encoding *hgl1* was expressed in three portions:

fragment C (nucleotides 46-1833) included the cysteine- and tryptophan-rich region, the cysteine-free region, and 277 amino acids of the cysteine-rich domain, i.e. amino acid residues 2-596;

fragment A (nucleotides 1492-3460) encoded the majority of the cysteine-rich domain, i.e. amino acid residues 482-1138;

fragment B (3461-3892) included 70 amino acids of the cysteine-rich domain, the putative membrane-spanning region, and the cytoplasmic tail, i.e. amino acid residues 1139-1276.

See Fig. 2B.

Each of these three fragments was inserted in frame by ligation into pGEX2T or pGEX3X to obtain these proteins as GST fusions. A diagram of the vectors constructed is shown in Figure 2A.

Fragment C was produced by PCR amplification. Primers were designed so that a *Bam*HI site was added to the 5′ end and an *Eco*RI site was added to the 3′ end

during the PCR process. The PCR product, fragment C, was then digested with restriction enzymes *Bam*HI and *Eco*RI, purified, and ligated into similarly digested pGEX3X. Fragments A and B were produced by digestion with *Eco*RI from plasmid clones (Mann, BJ et al. Proc Natl Acad Sci USA (1991) 88:3248-3252) and ligated into pGEX2T that had been digested with *Eco*RI. In the pGEX expression system a recombinant protein is expressed as a fusion protein with glutathione S-transferase (GST) from *Schistosoma japonicum* and is under the control of the *tac* promoter. The *tac* promoter is inducible by IPTG. The construction of the vectors and subsequent expression is further described in Mann, BJ et al. Infec and Immun (1993) 61:1772-1778, referenced above, and incorporated herein by reference.

Expression in the correct reading frame was verified for all constructs by sequencing and Western immunoblot analysis by testing for reactivity with anti-adhesion antisera (data not shown). Expression of the *hgl1* fusion proteins was shown to be inducible by IPTG. The GST protein produced from the original pGEX2T did not react with the anti-adhesion sera. The GST portion of the fusion protein has a molecular mass of 27.5 kD.

Example 2

Production of Recombinant Protein

The four vectors described above, as well as the host vector were transfected into competent *E. coli* hosts and expression of the genes encoding the fusion proteins was effected by induction with IPTG. Production of the fusion proteins was determined by Western blot SDS-PAGE analysis of the lysates.

Example 3

Reactivity of Recombinant 170 kD Subunit Fusion Proteins with MAbs

Induced cultures containing bacterial strains expressing *hgl1* fragment A, B, or C were harvested, lysed in sample buffer, and applied to an SDS-polyacrylamide gel. After electrophoresis, the proteins were transferred to Immobilon and incubated with anti-170-kD MAbs, specific for seven different epitopes. Characteristics of the individual MAbs are shown in Table 1. It will be noted that all the known epitopes are in the region of amino acids 596-1138.

TABLE 1. Characteristics of monoclonal antibodies directed against the galactose adhesion 170 kD subunit

Epitope #	Designation	Isotype[1]	Adherence[1]	Cytotoxicity[2]	C5b9 Resistance[3]	P[4]	NP[4]	Location[5]
1	3F4	IgG_1	Increases	Decreases	No effect	+	+	895-998
2	8A3	IgG_1	Increases	No effect	Decreases	+	+	895-998
3	7F4	IgG_{2b}	No effect	No effect	Decreases	+	-	1082-1138
4	8C12	IgG_1	Inhibits	Inhibits	Decreases	+	-	895-998
5	1G7	IgG_{2b}	Inhibits	Inhibits	Decreases	+	-	596-818
6	H85	IgG_{2b}	Inhibits[6]	Inhibits	Blocks	+	-	1033-1082
7	3D12	IgG_1	No effect	Not tested	Blocks	+		895-998

[1] Adherence was assayed by the binding of Chinese hamster ovary (CHO) cells to *E. histolytica* trophozoites and by binding of ^{125}I labeled purified colonic mucins to trophozoites. Petri, W.A. Jr., et al., J Immunol (1990) 144:4803-4809.

[2] The assay for cytotoxicity was CHO cell killing by *E. histolytica* trophozoites as measured by ^{51}Cr release from labeled CHO cells. Saffer, L.D., et al. Infect Immun (1991) 59:4681-4683.

[3] C5b9 resistance was assayed by the addition of purified complement components to *E. histolytica* trophozoites. The percent of amebic lysis was determined microscopically. Braga, L.L., et al. J Clin Invest (1992) 90:1131-1137.

[4] P and NP refer to reactivity of the MAb with pathogenic (P) and nonpathogenic (NP) species of *E. histolytica* as determined in an Elisa assay. Petri, W.A. Jr., et al. Infect Immun (1991) 58:1802-1806.

[5] Location of antibody binding site by amino acid number. Results presented herein.

[6] Inhibits adherence to CHO cells but not human colonic mucin glycoproteins. Petri, W.A. Jr., et al., J Immunol (1990) 144:4803-4809.

Fusion proteins B and C failed to react with any of the seven MAbs (data not shown). Fusion protein A, representing positions 482-1132, reacted with all seven MAbs representing all 7 epitopes and not a negative control developed with an irrelevant MAb, MOPC21. The MAbs were used at 10 μg/ml and polyclonal antibodies at 1:1000 dilution. These results indicated that these seven epitopes were contained within the 542 amino acids of the cysteine-rich extracellular domain of the 170 kD subunit.

The generation of 3′ deletions by controlled ExoIII digestion of fragment A of the 170 kD subunit is outlined in Figure 2B. Δ1 contains amino acid residues 482-1082; Δ2 contains amino acid residues 482-1032; Δ3 contains amino acid residues 482-998. The reactivities of the fusion proteins that include fragment A or either of two carboxy-terminal deletions (Δ3 and Δ4) with the seven distinct 170 kD-specific MAbs were determined. Deletion 3 reacted with MAb against epitopes 1-2, 4-5, and 7 but failed to react with MAbs recognizing epitopes 3 and 6; Deletion 4 which contains residues 498-894 reacted only with the MAb which recognizes epitope 5.

The five deletion derivatives of fusion protein A shown in Figure 2B, ranging in estimated size from 35 to 68 kD, were tested for reactivity to each MAb, and the reactivities of the deletions with each MAb are summarized in Fig. 3. The endpoints of the various deletions were determined by DNA sequencing with primers specific for the remaining *hgl1* sequence. MAbs recognizing epitopes 1 and 2, which increase amebic adherence to target cells, failed to react with recombinant lectin fusion proteins lacking amino acids 895 to 998. Similarly, MAbs recognizing epitope 4, an inhibitory epitope, and epitope 7, which has the effect of abrogating amebic lysis by complement, failed to react

with deletion mutants lacking this region. The MAb specific for epitope 6, which has inhibitory effects on amebic adherence and abrogates amebic lysis by complement, did not react with a recombinant protein missing amino acids 1033 to 1082. Recombinant proteins lacking amino acids 1082 to 1138 did not react with a MAb which is specific for the neutral epitope 3. Finally, a construct containing amino acids 482 to 818 was recognized only by the adherence-inhibitory epitope 5 MAb. The thus predicted locations of the MAb epitopes are listed in Table 1 above.

Example 4

Reactivity of 170 kD Fusion Proteins with Human Immune Sera

Since the galactose adhesion is a major target of the humoral immune response in the majority of immune individuals, the mapping of human B-cell epitopes of the 170 kD subunit was undertaken. The recombinant fusion proteins and ExoIII-generated deletion constructs of the 170 kD subunit were tested for reactivity with pooled human immune sera in the same manner as described for MAb reactivity. Nonimmune sera was used as a control. Fusion proteins A and C reacted with immune sera, whereas fusion protein B did not (data not shown). Human immune sera also reacted with deletion constructs Δ1, Δ2, and Δ3 but not with Δ4 or Δ10. Reactivity of immune sera with the different deletions localized major human B-cell epitopes to be within the first 482 amino acids and between amino acids 895 and 1138 (Fig. 3). This second region is the same area which contains six of the MAb epitopes. These results are consistent with a report by Zhang et al. supra, who found that sera from immune individuals reacted primarily with recombinant adhesion constructs containing amino acids 1 to 373 and 649 to 1202.

Thus, for use in assays to detect human antisera against *E. histolytica*, the useful epitope-bearing portions are as shown in Table 2.

Table 2

Positions	Epitope #	P/NP
2-482	?	?
1082-1138	3	P
1033-1082	6	P
895-998	1,2,4,7	both

The epitope-bearing portions indicated can be used alone, as fragments or as portions of chimeric or fusion proteins, or any combination of these epitope-bearing portions can be used.

Example 5

Immunization Using Recombinant Subunit Protein

A GST fusion protein with fragment A was prepared in *E. coli* as described in Example 1 above. This peptide contains an upstream GST derived peptide sequence followed by and fused to amino acids 432-1138 encoded by nucleotides 1492-3460 in Figure 1 herein. The protein is produced intracellularly; the cells were harvested and lysed and the lysates subjected to standard purification techniques to obtain the purified fusion protein.

Gerbils were immunized by intraperitoneal injection with 30 μg of purified fusion protein in complete Freund's adjuvant and then boosted at 2-4 weeks with 30 μg of the fusion protein in incomplete Freund's adjuvant.

The gerbils were challenged at 6 weeks by intrahepatic injection of $5x10^5$ amebic trophozoites and sacrificed 8 weeks later. The presence and size of amoebic liver abscesses was determined.

The results of the two experiments described above are shown in the tables below. The administration of the fusion protein reduced the size of abscesses in a statistically significant manner.

In experiment 1, six animals were used as controls and nine were administered the fusion protein; in experiment 2, seven animals were used as controls and seven were provided the fusion protein.

	Experiment 1		Experiment 2	
	Abscess Weight	% with Abscess	Abscess Weight	% with Abscess
Control	1.44±1.64	71%	4.76±1.78	100%
GST - (482-1138)	0.81±0.10*	100%	2.35±1.99	100%

* P<0.03 compared to control.
+ P<0.24 compared to control.

Example 6

Analysis of the Gene Family Encoding the 170 kD Subunit of *E. histolytica* Gal/GalNAc Adherence Lectin

This Example shows that the adhesin 170 kDa subunit of HM-1:IMSS strain *E. histolytica* is encoded by a gene family that includes *hgl1*, *hgl2* and a previously undescribed third gene herein designated *hgl3*. Since *hgl1* and *hgl2* were originally sequenced, in part, from different cDNA libraries, it was possible that they represented strain differences of a single gene. However, in this report both 5' and 3' termini of *hgl1*, *hgl2*, and *hgl3* were isolated and sequenced from the same lambda genomic library demonstrating unambiguously that *hgl* is a gene family.

Comparison of the amino acid sequences of the three heavy subunit genes showed that *hgl1* and *hgl2* are 89.2% identical, *hgl1* and *hgl3* are 95.2% identical, and *hgl2* and *hgl3* are 89.4% identical. Sequence variation within the gene family, however, appears to be nonrandomly distributed within the coding sequence. The

majority of the nonconservative amino acid substitutions as well as insertions and deletions occur in the amino third of the molecule. Comparison of the amino acid sequences of *hgl2* and *hgl3* reveal that 11 of the 19 nonconservative amino acid substitutions and 11 of the 13 residues inserted or deleted reside within the first 400 amino acid residues. A similar pattern of variation is present when *hgl1* and *hgl2* are compared. While *hgl1* and *hgl3* contain only two nonconservative substitutions, both are found within the first 400 residues although the 57 conservative substitutions appear to be more randomly distributed throughout the coding sequence. The high degree of sequence conservation between *hgl3* and *hgl1* suggest that they may have arisen from a recent gene duplication event.

All 97 cysteine residues were maintained in the three heavy subunit genes. The *hgl2* gene was originally reported lacking a single cysteine present in both *hgl1* and *hgl3*. However, this discrepancy has since been recognized as a sequencing error (Dr. E. Tannich, Bernhard Nocht Institute, Hamburg, Germany). The cysteine residues are nonrandomly distributed throughout the gene (Figure 4) with the highest concentration within the cysteine-rich domain between amino acid residues 379-1210. All seven identified epitopes recognized by murine monoclonal antibodies map to this region (Mann, B.J. *et al.* Infect Immun (1993) 61:1772-1778). As these monoclonal antibodies can block target cell adhesion, target cell lysis (Saffer, L.D. *et al.* Infect Immun (1991) 59:4681-4683), and/or resistance to host complement-mediated lysis (Braga, L.L. *et al.* J Clin Invest (1992) 90:1131-1137), the conservation of cysteine residues may play an important role in maintaining the conformation of this important region of *hgl*.

A minimum of three genes are shown to make up the heavy subunit gene family. While it is not possible

to rule out the existence of additional *hgl* genes, the Southern blot and library screen data can be explained by a gene family of three members. As the genomic library was screened separately with a 5′ and a 3′ *hgl* specific probe, additional heavy subunit genes would be isolated even if they contained only partial identity with the gene family at only one end or even if one termini of an additional gene had been lost during library amplification. The library screen looked at more than 3.2×10^8 bases of genomic DNA in an organism with an estimated genome size of $10^{7.5}$ bases (Gelderman, A.H. *et al*. J Parasitol (1971) 57:906-911). Thus, a full genomic equivalent was screened at low stringency for genes containing identity at either end.

The Northern data indicated that all three genes were expressed in the amebae. As the messages of *hgl*1-3 are predicted to comigrate at 4.0 kb, differential hybridization was required to ascertain expression of individual genes. Due to the high degree of identity between *hgl*1-3, relatively short oligonucleotides (17-21 bases) were synthesized specific for regions where the three genes diverge. Each probe was compared by computer analysis to the other *hgl* genes to be certain that they were sufficiently divergent to prevent cross hybridization. Hybridization and wash conditions were highly stringent for such A/T rich probes and were done at temperatures 5°C or less below the predicted Tm based upon nearest neighbor analysis. While it is impossible to rule out cross hybridization with other *hgl* gene members, these precautions make such an event less likely.

The Northern blot also indicates that abundance of mRNA for all three genes decreased as the amebae progressed from log to stationary growth. This finding correlates with data which indicate that late log and stationary phase amebae have a decreased ability to

adhere to, lyse, and phagocytose target cells (Orozco, E. *et al.* (1988) "The role of phagocytosis in the pathogenic mechanism of *Entamoeba histolytica*. In: Amebiasis: Human infection by *Entamoeba histolytica* (Ravdin J.I., ed), pp. 326-338. John Wiley & Sons, Inc., New York.

Details of the experimental methods and results of the characterization of the *hgl* multigene family are presented below.

Library Screen. A lambda Zap® II library containing randomly sheared 4-5 kb fragments of genomic DNA from HM-1:IMSS strain *E. histolytica* was kindly provided by Dr. J. Samuelson at Harvard University (Kumar, A. *et al.* Proc Natl Acad Sci USA (1992) 89:10188-10192). Over 80,000 plaques from the library were screened on a lawn of XL-1 Blue *E. coli* (Strategene, La Jolla, CA). Duplicate plaque lifts, using Hybond-N membranes (Amersham, Arlington Heights, IL), were placed in a prehybridization solution consisting of 6x SSC (.89 M sodium chloride and 90 mM sodium citrate), 5x Denhardts solution, .5% SDS, 50 mM $NaPO_4$ (pH 6.7), and 100 μg/ml salmon sperm DNA for a minimum of 4 hours at 55°C. A 5' and 3' DNA fragment of *hgl*1 (nucleotides 106-1946 and 3522-3940 respectively) were [α-^{32}P]dCTP (Amersham) labeled using the Random Primed DNA labeling Kit according to the manufacturer's instructions (Boehringer Mannheim, Mannheim, Germany) and hybridized separately to the membranes overnight at 55°C in prehybridization solution. Membranes were rinsed once and washed once for 15 minutes at room temperature in 2x SSC, .1% SDS, then washed once for 15 minutes at room temperature, and twice at 55°C for 20 minutes in .1x SSC, .1% SDS. Plaques that hybridized with the 5' or the 3 radiolabeled probe on both duplicate filters were isolated and purified.

Northern blot and hybridization. Total RNA was harvested from amebae using the guanidinium isothiocyanate method (RNagen, Promega, Madison, WI).

Polyadenylated RNA was purified from total RNA using PolyATract System 1000 (Promega). RNA was electrophoresed through a formaldehyde gel and transferred to a nylon Zetabind membrane (Cuno) using 25 mM phosphate buffer (pH 7.5) as described (Sambrook, J. *et al.* (1989) Molecular Cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). The membrane was incubated in prehybridization solution and incubated at 37°C for at least two hours. Oligonucleotides (18-22 nucleotides long) were end-labeled using polynucleotide kinase and [γ-P^{32}]ATP (Sambrook, J. *et al.* (1989) Molecular Cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), added to the hybridization mixture and the membrane, and incubated at 37°C overnight. The membrane was then washed once at room temperature for 10 minutes, once at 37°C for 10 minutes; and twice at 40-44°C for 15 minutes each in 2x SSC, .1% SDS. The radiolabeled probes used were: 5'-TTTGTCACTATTTTCTAC-3', *hgl1*; 5'-TATCTCCATTTGGTTGA-3', *hgl2*; 5'-TTTGTCACTATTTTCTAC-3', *hgl3*; and 5'-CCCAAGCATATTTGAATG-3', EF-1α (Plaimauer, B. *et al.* DNA Cell Biol (1993) 12:89-96).

Characterization of the *hgl3* gene. The *hgl3* open reading frame was 3876 bases and would result in a predicted translation product of 1292 amino acids (Figure 4). The predicted translation products of *hgl1* and *hgl2* would be 1291 and 1285 amino acids respectively. A putative signal sequence and a transmembrane domain were identified in the amino acid sequence of *hgl3* similar to *hgl1* and *hgl2*. The amino-terminal amino acid sequence of the mature *hgl3* protein, determined by Edman degradation (Mann, B.J. *et al.* Proc Natl Acad Sci USA (1991) 88:3248-3252), was assigned residue number 1. Previous analysis of *hgl1* and *hgl2* identified a large, conserved, extracellular region which was 11% cysteine, designated

the cysteine-rich domain (Mann, B.J. *et al.* Parasit Today (1991) 7:173-176) (Fig. 2). Sequence analysis of *hgl3* revealed that all 97 cysteine residues present within this region were also conserved in both of the previously reported heavy subunit genes.

A schematic comparison (Figure 5) of heavy subunit gene sequences revealed a high degree of amino acid sequence identity. However, seven sites, ranging from 3-24 nucleotides, were found where an insertion or deletion had occurred in one subunit relative to another, all of which maintained the open reading frame. Both *hgl1* and *hgl3* contained a large number of nonconservative amino acid substitutions when compared to *hgl2*, making them 89.2% and 89.4% identical to *hgl2* respectively. While the comparison of *hgl1* and *hgl3* revealed only two nonconservative substitutions, 57 conservative amino acid substitutions and 3 single residue insertion/deletions making them 95.2% identical.

All 16 potential sites of glycosylation present in *hgl1* were conserved in *hgl3*. A sequence analysis of *hgl2* indicated that it contained only 9 such sites, although all 9 were present in *hgl1* and *hgl3*. Glycosylation appears to account for approximately 6% of the heavy subunits' apparent molecular mass (Mann, B.J. *et al.* Proc Natl Acad Sci USA (1991) 88:3248-3252).

All three heavy subunits are expressed. Since *hgl3* was isolated from a genomic library, it was unknown if this gene was transcribed. Polyadenylated RNA was harvested from amebae in both log and stationary phase growth. Probes specific for *hgl1*, *hgl2*, or *hgl3* were hybridized to a Northern blot and identified an RNA band of the predicted size of 4.0 kb.

As the messages of *hgl1*-3 are predicted to comigrate at 4.0 kb, differential hybridization was required to ascertain expression of individual genes using Northern blots. Due to the high degree of identity

between *hgl*1-3, relatively short oligonucleotides (17-21 bases) were synthesized specific for regions where the three genes diverge. Each probe was compared by computer analysis to the other *hgl* genes to be certain that they were sufficiently divergent to prevent cross hybridization. Hybridization and wash conditions were highly stringent for such A/T rich probes and were done at temperatures 5°C or less below the predicted Tm based upon nearest neighbor analysis. While it is impossible to rule out cross hybridization with other *hgl* gene members, these precautions make such an event less likely.

The message abundance decreased significantly as the amebic trophozoites passed from log phase growth (lane A) to stationary phase growth (lane B) while the control gene, EF-1α, either remained constant or increased slightly. This finding correlates with data indicating that late log and stationary phase amebae have a decreased ability to adhere to, lyse, and phagocytose target cells (Orozco, E. *et al*. (1988) "The role of phagocytosis in the pathogenic mechanism of *Entamoeba histolytica*. In: Amebiasis: Human infection by *Entamoeba histolytica* (Ravdin J.I., ed), pp. 326-338. John Wiley & Sons, Inc., New York.

Estimation of the number of heavy subunit genes. The observations herein confirm that the adhesin 170 kDa subunit of HM-1:IMSS strain *E. histolytica* is encoded by a gene family that includes *hgl*1, *hgl*2 and a previously undescribed third gene which is designated *hgl*3. Since *hgl*1 and *hgl*2 were originally sequenced, in part, from different cDNA libraries, it was possible that they represented strain differences of a single gene. However, in the present work both 5′ and 3′ termini of *hgl*1, *hgl*2, and *hgl*3 were isolated and sequenced from the same lambda genomic library, demonstrating unambiguously that *hgl* is a gene family.

Comparison of the amino acid sequences of the three heavy subunit genes found that *hgl1* and *hgl2* are 89.2% identical, *hgl1* and *hgl3* are 95.2% identical, and *hgl2* and *hgl3* are 89.4% identical. Sequence variation within the gene family, however, appears to be nonrandomly distributed within the coding sequence. The majority of the nonconservative amino acid substitutions as well as insertions and deletions occur in the amino third of the molecule. Comparison of the amino acid sequences of *hgl2* and *hgl3* reveal that 11 of the 19 nonconservative amino acid substitutions and 11 of the 13 residues inserted or deleted reside within the first 400 amino acid residues. A similar pattern of variation is present when *hgl1* and *hgl2* are compared. While *hgl1* and *hgl3* contain only two nonconservative substitutions, both are found within the first 400 residues although the 57 conservative substitutions appear to be more randomly distributed throughout the coding sequence. The high degree of sequence conservation between *hgl3* and *hgl1* suggest that they may have arisen from a recent gene duplication event.

All 97 cysteine residues were maintained in the three heavy subunit genes. The *hgl2* gene was originally reported lacking a single cysteine present in both *hgl1* and *hgl3*. However, this discrepancy has since been recognized as a sequencing error (Dr. E. Tannich, Bernhard Nocht Institute, Hamburg, Germany, personal communication). The cysteine residues are nonrandomly distributed throughout the gene (Fig. 1) with the highest concentration within the cysteine-rich domain between amino acid residues 379-1210. All seven identified epitopes recognized by murine monoclonal antibodies map to this region (Mann, B.J. *et al.* Infect Immun (1993) 61:1772-1778). As these monoclonal antibodies can block target cell adhesion, target cell lysis (Saffer, L.D. *et al.* Infect Immun (1991) 59:4681-4683), and/or resistance

to host complement-mediated lysis (Braga, L.L. *et al*. J Clin Invest (1992) 90:1131-1137), the conservation of cysteine residues may play an important role in maintaining the conformation of this important region of *hgl*.

A minimum of three genes have been shown to make up the heavy subunit gene family, as described herein. While it is not possible to rule out the existence of additional *hgl* genes, Southern blot analyses and library screen data can best be explained by a gene family of three members. For Southern blots, two restriction enzymes were identified, *Dde*I and *Hin*dIII, that cut genomic DNA to completion and resulted in analyzable restriction fragments. As the membrane was hybridized with a fragment of *hgl*1 corresponding to nucleotides 1556 to 3522, two bands of >976 and 1965 nucleotides should have been present from *hgl*3. This central *hgl*1 radioprobe would hybridize with three bands of 1158, 810 and >1080 nucleotides from *hgl*1 and would hyribidze with five bands of 819, 312, 55, 755, and >1080 nucleotides from *hgl*2. The Southern blot showed 7 bands for genomic DNA disgested with *Dde*I, at 4200, 3700, 2100, 1800, 1300, 840, and 760 nucleotides. As the 819 and 810 nucleotide bands would be expected to comigrate, all the bands observed with *Dde*I digestion are explained by the restriction maps of *hgl*1-3.

*Hin*dIII has no restriction sites in *hgl*1-3 within the coding region and would result in each gene being represented by a single band greater than 4.0 kb. The Southern blot showed three bands at 17500, 5600, and 4200 nucleotides. Should an additional heavy subunit gene exist, its *Dde*I and *Hin*dIII fragments would need to comigrate with *hgl*1-3 bands, be so divergent that they failed to hybridize with the *hgl*1 probe under very low stringency, or be too large to be resolved and transferred.

As to the genomic screening data, the genomic library was screened separately with a 5' and a 3' *hgl* specific probe, additional heavy subunit genes would be isolated even if they contained only partial identity with the gene family at only one end or even if one termini of an additional gene had been lost during library amplification. The library screen looked at more than 3.2×10^8 bases of genomic DNA in an organism with an estimated genome size of $10^{7.5}$ bases (Gelderman, A.H. *et al*. J Parasitol (1971) 57:906-911). Thus, a full genomic equivalent was screened at low stringency for genes containing identity at either end. Of 7 clones identified with the 5' heavy subunit-specific probe, 4 contained inserts that matched the reported sequence for *hgl*1, 2 matched the sequence of *hgl*2, and 1 clone represented *hgl*3. Of eight clones obtained using the 3' radiolabeled fragment, 1 matched the sequence for *hgl*1, 5 matched the sequence of *hgl*2, and 2 represented *hgl*3. No termini were found that did not match the sequence of *hgl*1, *hgl*2 or *hgl*3.

CLAIMS

1. A method to detect *Entamoeba histolytica* antibodies in a biological sample which method comprises contacting said sample with an epitope-bearing portion of the 170 kD subunit of *E. histolytica* Gal/GalNAc adherence lectin under conditions wherein said portion forms a complex with any antibodies immunoreactive with said epitope present in said sample, and

assessing the presence, absence or amount of said complex,

wherein said epitope-bearing portion is nonglycosylated and is in a form obtained by recombinant production in a procaryotic host cell culture.

2. The method of claim 1 wherein said contacting is conducted by providing said epitope-bearing portion coupled to a solid support and treating said solid support with the biological sample.

3. The method of claim 2 wherein said contacting and assessing steps are conducted in a Western blot procedure.

4. The method of claim 2 which further comprises treating said solid support with a specific binding partner for said epitope under conditions wherein any antibody to said epitope in the biological sample competes with said specific binding partner for said epitope-bearing portion.

5. The method of claim 4 wherein said specific binding partner contains a detectable label and said assessing is conducted by measuring the effect of the presence of biological sample on the amount of label retained on the solid support.

6. The method of claim 5 wherein said specific binding partner is an antibody or immunologically reactive portion thereof.

7. The method of claim 1 wherein said epitope-bearing portion contains a detectable label.

8. The method of claim 1 wherein said contacting is conducted by providing said biological sample coupled to a solid support and treating said support with said epitope-bearing portion.

9. The method of claim 1 wherein said epitope-bearing portion is characteristic of the pathogenic form of *E. histolytica*.

10. The method of claim 1 wherein said epitope-bearing portion is characteristic of the non-pathogenic form of *E. histolytica*.

11. The method of claim 1 wherein said epitope-bearing portion is characteristic of both the pathogenic and non-pathogenic form of *E. histolytica*.

12. The method of claim 1 wherein said epitope-bearing portion consists essentially of amino acids of said subunit as shown in Figure 1B in positions selected from the group consisting of 2-482, 1082-1138, 1033-1082, and 895-998, or the corresponding amino acids in a naturally occurring variant thereof.

13. A nonglycosylated epitope-bearing portion of the 170 kD subunit of *E. histolytica* Gal/GalNAc adherence lectin in a form produced by recombinant procaryotic cell culture.

14. The epitope-bearing portion of claim 13 which comprises the complete subunit.

15. The epitope-bearing portion of claim 13 which consists essentially of amino acids of said subunit as shown in Figure 1B in positions selected from the group consisting of 2-482, 1082-1138, 1033-1082, and 895-998, or the corresponding amino acids in a naturally occurring variant thereof.

16. An article of manufacture useful for the analysis of a biological sample for antibodies immunoreactive with *E. histolytica*, which article consists essentially of a solid support coupled to an epitope-bearing portion of the 170 kD subunit of *E. histolytica* Gal/GalNAc adherence lectin, which portion is nonglycosylated and which portion is in a form produced by recombinant procaryotic cell culture.

17. A kit for the analysis of biological sample for antibodies immunoreactive with *E. histolytica*, which kit comprises an epitope-bearing portion of the 170 kD subunit of *E. histolytica* Gal/GalNAc adherence lectin which is nonglycosylated and which is in a form produced by recombinant procaryotic cell culture, along with the reagents suitable for assessing the formation of a complex between any said antibody in said biological fluid and said epitope-bearing portion.

18. A method to immunize a subject against *Entamoeba histolytica* infection which method comprises administering to said subject an effective amount of an epitope-bearing portion of the 170 kD subunit of *E. histolytica* Gal/GalNAc adherence lectin which epitope-bearing portion is nonglycosylated and is in a form

obtained by recombinant production in a procaryotic host cell culture.

19. The method of claim 18 wherein said epitope-bearing portion consists essentially of amino acids of said subunit as shown in Figure 1B in positions selected from the group consisting of 482-1138, 596-1138, 895-998, 1033-1082 and 1082-1138 or the corresponding amino acids in a naturally occurring variant of said 170 kD subunit.

20. The method of claim 19 wherein said positions are 1082-1138 or 1033-1082.

21. The method of claim 19 wherein said positions are 482-1138 or 596-1138.

22. A vaccine for immunizing a subject against *E. histolytica* infection which vaccine comprises a nonglycosylated epitope-bearing portion of the 170 kD subunit of *E. histolytica* Gal/GalNac adherence lectin in a form produced by recombinant procaryotic cell culture.

23. The vaccine of claim 22 wherein the epitope-bearing portion consists essentially of amino acids of said subunit as shown in Figure 1B in positions selected from the group consisting of 895-998, 1033-1082 and 1082-1138 or the corresponding amino acids in a naturally occurring variant of said 170 kD subunit.

24. The vaccine of claim 22 wherein said positions are 1082-1138 or 1033-1082.

25. The vaccine of claim 22 wherein said positions are 482-1138 or 596-1138.

26. The epitope-bearing portion of claim 13 which consists essentially of amino acids of said subunit as shown in Figure 1B in positions 482-1138, or the corresponding amino acids in a naturally occurring variant thereof.

27. The method of claim 19 wherein said positions are 482-1138.

28. The vaccine of claim 22 wherein said positions are 482-1138.

29. The method of claim 1 wherein said epitope-bearing portion is selected from the group consisting of: a portion which is characteristic of the subunit encoded by the *hgl1* gene, a portion which is characteristic of the subunit encoded by the *hgl2* gene, a portion which is characteristic of the subunit encoded by the *hgl3* gene, and a portion which is shared by the subunit encoded by the *hgl1*, *hgl2* and *hgl3* genes.

```
  1 ATG AAA TTA TTA TTA TTA AAT ATC TTA TTA TTA TGT TGT CTT  42
    M   K   L   L   L   L   N   I   L   L   L   C   C   L

 43 GCA GAT AAA CTT GAT GAA TTT TCA GCA GAT AAT GAC TAT TAT  84
    A   D   K   L   D   E   F   S   A   D   N   D   Y   Y

 85 GAC GGT GGT ATT ATG TCT CGT GGA AAG AAT GCA GGT TCA TGG 126
    D   G   G   I   M   S   R   G   K   N   A   G   S   W

127 TAT CAT TCT TAC ACT CAC CAA TAT GAT GTT TTC TAT TAT TTA 168
    Y   H   S   Y   T   H   Q   Y   D   V   F   Y   Y   L

169 GCT ATG CAA CCA TGG AGA CAT TTT GTA TGG ACT ACA TGC GAT 210
    A   M   Q   P   W   R   H   F   V   W   T   T   C   D

211 AAA AAT GAT AAT ACA GAA TGT TAT AAA TAT ACT ATC AAT GAA 252
    K   N   D   N   T   E   C   Y   K   Y   T   I   N   E

253 GAT CAT AAT GTA AAG GTT GAA GAT ATT AAT AAA ACA AAT ATT 294
    D   H   N   V   K   V   E   D   I   N   K   T   N   I

295 AAA CAA GAT TTT TGT CAA AAA GAA TAT GCA TAT CCA ATT GAA 336
    K   Q   D   F   C   Q   K   E   Y   A   Y   P   I   E

337 AAA TAT GAA GTT GAT TGG GAC AAT GTT CCA GTT GAT GAA CAA 378
    K   Y   E   V   D   W   D   N   V   P   V   D   E   Q

379 CGA ATT GAA AGT GTA GAT ATT AAT GGA AAA ACT TGT TTT AAA 420
    R   I   E   S   V   D   I   N   G   K   T   C   F   K

421 TAT GCA GCT AAA AGA CCA TTG GCT TAT GTT TAT TTA AAT ACA 462
    Y   A   A   K   R   P   L   A   Y   V   Y   L   N   T

463 AAA ATG ACA TAT GCA ACA AAA ACT GAA GCA TAT GAT GTT TGT 504
    K   M   T   Y   A   T   K   T   E   A   Y   D   V   C

505 AGA ATG GAT TTC ATT GGA GGA AGA TCA ATT ACA TTC AGA TCA 546
    R   M   D   F   I   G   G   R   S   I   T   F   R   S

547 TTT AAC ACA GAG AAT AAA GCA TTT ATT GAT CAA TAT AAT ACA 588
    F   N   T   E   N   K   A   F   I   D   Q   Y   N   T

589 AAC ACT ACA TCA AAA TGT CTT CTT AAT GTA TAT GAT AAT AAT 630
    N   T   T   S   K   C   L   L   N   V   Y   D   N   N

631 GTT AAT ACA CAT CTT GCA ATT ATC TTT GGT ATT ACT GAT TCT 672
    V   N   T   H   L   A   I   I   F   G   I   T   D   S

673 ACA GTC ATT AAA TCA CTT CAA GAG AAT TTA TCT CTT TTA AGT 714
    T   V   I   K   S   L   Q   E   N   L   S   L   L   S

715 CAA CTA AAA ACA GTC AAA GGA GTA ACA CTC TAC TAT CTT AAA 756
    Q   L   K   T   V   K   G   V   T   L   Y   Y   L   K

757 GAT GAT ACT TAT TTT ACA GTT AAT ATT ACT TTA GAT CAA TTA 798
```

FIG. 1A-1

```
      D   D   T   Y   F   T   V   N   I   T   L   D   Q   L
799 AAA TAT GAT ACA CTT GTC AAA TAC ACA GCA GGA ACA GGA CAA 840
      K   Y   D   T   L   V   K   Y   T   A   G   T   G   Q
841 GTT GAT CCA CTT ATT AAT ATT GCT AAG AAT GAT TTA GCT ACT 882
      V   D   P   L   I   N   I   A   K   N   D   L   A   T
883 AAA GTT GCA GAT AAA AGT AAA GAT AAA AAT GCA AAT GAT AAA 924
      K   V   A   D   K   S   K   D   K   N   A   N   D   K
925 ATC AAA AGA GGA ACT ATG ATT GTG TTA ATG GAT ACT GCA CTT 966
      I   K   R   G   T   M   I   V   L   M   D   T   A   L
967 GGA TCA GAA TTT AAT GCA GAA ACA GAA TTT GAT AGA AAG AAT 1008
      G   S   E   F   N   A   E   T   E   F   D   R   K   N
1009 ATT TCA GTT CAT ACT GTT GTT CTT AAT AGA AAT AAA GAC CCA 1050
      I   S   V   H   T   V   V   L   N   R   N   K   D   P
1051 AAG ATT ACA CGT AGT GCA TTG AGA CTT GTT TCA CTT GGA CCA 1092
      K   I   T   R   S   A   L   R   L   V   S   L   G   P
1093 CAT TAT CAT GAA TTT ACA GGT AAT GAT GAA GTT AAT GCA ACA 1134
      H   Y   H   E   F   T   G   N   D   E   V   N   A   T
1135 ATC ACT GCA CTT TTC AAA GGA ATT AGA GCC AAT TTA ACA GAA 1176
      I   T   A   L   F   K   G   I   R   A   N   L   T   E
1177 AGA TGT GAT AGA GAT AAA TGT TCA GGA TTT TGT GAT GCA ATG 1218
      R   C   D   R   D   K   C   S   G   F   C   D   A   M
1219 AAT AGA TGC ACA TGT CCA ATG TGT TGT GAG AAT GAT TGT TTC 1260
      N   R   C   T   C   P   M   C   C   E   N   D   C   F
1261 TAT ACA TCC TGT GAT GTA GAA ACA GGA TCA TGT ATT CCA TGG 1302
      Y   T   S   C   D   V   E   T   G   S   C   I   P   W
1303 CCT AAA GCT AAA CCA AAA GCA AAG AAA GAA TGT CCA GCA ACA 1344
      P   K   A   K   P   K   A   K   K   E   C   P   A   T
1345 TGT GTA GGC TCA TAT GAA TGT AGA GAT CTT GAA GGA TGT GTT 1386
      C   V   G   S   Y   E   C   R   D   L   E   G   C   V
1387 GTT ACA AAA TAT AAT GAC ACA TGC CAA CCA AAA GTG AAA TGC 1428
      V   T   K   Y   N   D   T   C   Q   P   K   V   K   C
1429 ATG GTA CCA TAT TGT GAT AAT GAT AAG AAT CTA ACT GAA GTA 1470
      M   V   P   Y   C   D   N   D   K   N   L   T   E   V
1471 TGT AAA CAA AAA GCT AAT TGT GAA GCA GAT CAA AAA CCA AGT 1512
      C   K   Q   K   A   N   C   E   A   D   Q   K   P   S
1513 TCT GAT GGA TAT TGT TGG AGT TAT ACA TGT GAC CAA ACT ACT 1554
      S   D   G   Y   C   W   S   Y   T   C   D   Q   T   T
1555 GGT TTT TGT AAG AAA GAT AAA CGA GGT AAA GAA ATG TGT ACA 1596
```

FIG. 1A-2

```
      G   F   C   K   K   D   K   R   G   K   E   M   C   T
1597 GGA AAG ACA AAT AAT TGT CAA GAA TAT GTT TGT GAT TCA GAA 1638
      G   K   T   N   N   C   Q   E   Y   V   C   D   S   E

1639 CAA AGA TGT AGT GTT AGA GAT AAA GTA TGT GTA AAA ACA TCA 1680
      Q   R   C   S   V   R   D   K   V   C   V   K   T   S

1681 CCA TAC ATT GAA ATG TCA TGT TAT GTA GCC AAG TGT AAT CTC 1722
      P   Y   I   E   M   S   C   Y   V   A   K   C   N   L

1723 AAT ACA GGT ATG TGT GAG AAC AGA TTA TCA TGT GAT ACA TAC 1764
      N   T   G   M   C   E   N   R   L   S   C   D   T   Y

1765 TCA TCA TGT GGT GGA GAT TCT ACA GGA TCA GTA TGT AAA TGT 1806
      S   S   C   G   G   D   S   T   G   S   V   C   K   C

1807 GAT TCT ACA ACT GGT AAT AAA TGT CAA TGT AAT AAA GTA AAA 1848
      D   S   T   T   G   N   K   C   Q   C   N   K   V   K

1849 AAT GGT AAT TAT TGT AAT TCT AAA AAC CAT GAA ATT TGT GAT 1890
      N   G   N   Y   C   N   S   K   N   H   E   I   C   D

1891 TAT ACA GGA ACA ACA CCA CAA TGT AAA GTG TCT AAT TGT ACA 1932
      Y   T   G   T   T   P   Q   C   K   V   S   N   C   T

1933 GAA GAT CTT GTT AGA GAT GGA TGT CTT ATT AAG AGA TGC AAT 1974
      E   D   L   V   R   D   G   C   L   I   K   R   C   N

1975 GAA ACA AGT AAA ACA ACA TAT TGG GAG AAT GTT GAT TGT TCA 2016
      E   T   S   K   T   T   Y   W   E   N   V   D   C   S

2017 AAC ACT AAG ATT GAA TTT GCT AAA GAT GAT AAA TCT GAA ACT 2058
      N   T   K   I   E   F   A   K   D   D   K   S   E   T

2059 ATG TGT AAA CAA TAT TAT TCA ACT ACA TGT TTG AAT GGA AAA 2100
      M   C   K   Q   Y   Y   S   T   T   C   L   N   G   K

2101 TGT GTT GTT CAA GCA GTT GGT GAT GTT TCT AAT GTA GGA TGT 2142
      C   V   V   Q   A   V   G   D   V   S   N   V   G   C

2143 GGA TAT TGT TCA ATG GGA ACA GAT AAT ATT ATT ACA TAT CAT 2184
      G   Y   C   S   M   G   T   D   N   I   I   T   Y   H
```

FIG. 1A-3

```
2185 GAT GAT TGT AAT TCA CGT AAA TCA CAA TGT GGA AAC TTT AAT 2226
       D   D   C   N   S   R   K   S   Q   C   G   N   F   N

2227 GGT AAA TGT ATT AAA GGC AGT GAC AAT TCT TAT TCT TGT GTA 2268
       G   K   C   I   K   G   S   D   N   S   Y   S   C   V

2269 TTT GAA AAA GAT AAA ACT TCT TCT AAA TCA GAT AAT GAT ATT 2310
       F   E   K   D   K   T   S   S   K   S   D   N   D   I

2311 TGT GCT GAA TGT TCT AGT TTA ACA TGT CCA GCT GAT ACT ACA 2352
       C   A   E   C   S   S   L   T   C   P   A   D   T   T

2353 TAC AGA ACA TAT ACA TAT GAC TCA AAA ACA GGA ACA TGT AAA 2394
       Y   R   T   Y   T   Y   D   S   K   T   G   T   C   K

2395 GCA ACT GTT CAA CCA ACA CCA GCA TGT TCA GTA TGT GAA AGT 2436
       A   T   V   Q   P   T   P   A   C   S   V   C   E   S

2437 GGT AAA TTT GTA GAG AAA TGC AAA GAT CAA AAA TTA GAA CGT 2478
       G   K   F   V   E   K   C   K   D   Q   K   L   E   R

2479 AAA GTC ACT TTA GAA AAT GGA AAA GAA TAT AAA TAC ACC ATT 2520
       K   V   T   L   E   N   G   K   E   Y   K   Y   T   I

2521 CCA AAA GAT TGT GTC AAT GAA CAA TGC ATT CCA AGA ACA TAC 2562
       P   K   D   C   V   N   E   Q   C   I   P   R   T   Y

2563 ATA GAT TGT TTA GGT AAT GAT GAT AAC TTT AAA TCT ATT TAT 2604
       I   D   C   L   G   N   D   D   N   F   K   S   I   Y

2605 AAC TTC TAT TTA CCA TGT CAA GCA TAT GTT ACA GCT ACC TAT 2646
       N   F   Y   L   P   C   Q   A   Y   V   T   A   T   Y

2647 CAT TAC AGT TCA TTA TTC AAT TTA ACT AGT TAT AAA CTT CAC 2688
       H   Y   S   S   L   F   N   L   T   S   Y   K   L   H

2689 TTA CCA CAA AGT GAA GAA TTT ATG AAA GAG GCA GAC AAA GAA 2730
       L   P   Q   S   E   E   F   M   K   E   A   D   K   E

2731 GCA TAT TGT ACA TAC GAA ATA ACA ACA AGA GAA TGT AAA ACA 2772
       A   Y   C   T   Y   E   I   T   T   R   E   C   K   T

2773 TGT TCA TTA ATT GAA ACT AGA GAA AAA GTC CAA GAA GTT GAT 2814
       C   S   L   I   E   T   R   E   K   V   Q   E   V   D

2815 TTG TGT GCA GAA GAA ACT AAG AAT GGA GGA GTT CCA TTC AAA 2856
       L   C   A   E   E   T   K   N   G   G   V   P   F   K

2857 TGT AAG AAT AAC AAT TGC ATT ATT GAT CCT AAC TTT GAT TGT 2898
       C   K   N   N   N   C   I   I   D   P   N   F   D   C

2899 CAA CCT ATT GAA TGT AAG ATT CAA GAG ATT GTT ATT ACA GAA 2940
       Q   P   I   E   C   K   I   Q   E   I   V   I   T   E

2941 AAA GAT GGA ATA AAA ACA ACA ACA TGT AAA AAT ACT ACA AAA 2982
       K   D   G   I   K   T   T   T   C   K   N   T   T   K
```

FIG. 1A-4

```
2983 GCA ACA TGT GAC ACT AAC AAT AAG AGA ATA GAA GAT GCA CGT 3024
       A   T   C   D   T   N   N   K   R   I   E   D   A   R

3025 AAA GCA TTC ATT GAA GGA AAA GAA GGA ATT GAG CAA GTA GAA 3066
       K   A   F   I   E   G   K   E   G   I   E   Q   V   E

3067 TGT GCA AGT ACT GTT TGT CAA AAT GAT AAT AGT TGT CCA ATT 3108
       C   A   S   T   V   C   Q   N   D   N   S   C   P   I

3109 ATT ACT GAT GTA GAA AAA TGT AAT CAA AAC ACA GAA GTA GAT 3150
       I   T   D   V   E   K   C   N   Q   N   T   E   V   D

3151 TAT GGA TGT AAA GCA ATG ACA GGA GAA TGT GAT GGT ACT ACA 3192
       Y   G   C   K   A   M   T   G   E   C   D   G   T   T

3193 TAT CTT TGT AAA TTT GTA CAA CTT ACT GAT GAT CCA TCA TTA 3234
       Y   L   C   K   F   V   Q   L   T   D   D   P   S   L

3235 GAT AGT GAA CAT TTT AGA ACT AAA TCA GGA GTT GAA CTT AAC 3276
       D   S   E   H   F   R   T   K   S   G   V   E   L   N

3277 AAT GCA TGT TTG AAA TAT AAA TGT GTT GAG AGT AAA GGA AGT 3318
       N   A   C   L   K   Y   K   C   V   E   S   K   G   S

3319 GAT GGA AAA ATC ACA CAT AAA TGG GAA ATT GAT ACA GAA CGA 3360
       D   G   K   I   T   H   K   W   E   I   D   T   E   R

3361 TCA AAT GCT AAT CCA AAA CCA AGA AAT CCA TGC GAA ACC GCA 3402
       S   N   A   N   P   K   P   R   N   P   C   E   T   A

3403 ACA TGT AAT CAA ACA ACT GGA GAA ACT ATT TAC ACA AAG AAA 3444
       T   C   N   Q   T   T   G   E   T   I   Y   T   K   K

3445 ACA TGT ACT GTT TCA GAA TTC CCA ACA ATC ACA CCA AAT CAA 3486
       T   C   T   V   S   E   F   P   T   I   T   P   N   Q

3487 GGA AGA TGT TTC TAT TGT CAA TGT TCA TAT CTT GAC GGT TCA 3528
       G   R   C   F   Y   C   Q   C   S   Y   L   D   G   S

3529 TCA GTT CTT ACT ATG TAT GGA GAA ACA GAT AAA GAA TAT TAT 3570
       S   V   L   T   M   Y   G   E   T   D   K   E   Y   Y
```

FIG. 1A-5

```
3571 GAT CTT GAT GCA TGT GGT AAT TGT CGT GTT TGG AAT CAG ACA 3612
      D   L   D   A   C   G   N   C   R   V   W   N   Q   T

3613 GAT AGA ACA CAA CAA CTT AAT AAT CAC ACC GAG TGT ATT CTC 3654
      D   R   T   Q   Q   L   N   N   H   T   E   C   I   L

3655 GCA GGA GAA ATT AAT AAT GTT GGA GCT ATT GCA GCG GCA ACT 3696
      A   G   E   I   N   N   V   G   A   I   A   A   A   T

3697 ACT GTG GCT GCT GTT ATA GTT GCA GTT GTA GTT GCA TTA ATT 3738
      T   V   A   A   V   I   V   A   V   V   V   A   L   I

3739 GTT GTT TCT ATT GGA TTA TTT AAG ACT TAT CAA CTT GTT TCA 3780
      V   V   S   I   G   L   F   K   T   Y   Q   L   V   S

3781 TCA GCT ATG AAG AAT GCC ATT ACA ATA ACT AAT GAA AAT GCA 3822
      S   A   M   K   N   A   I   T   I   T   N   E   N   A

3823 GAA TAT GTT GGA GCA GAT AAT GAA GCA ACT AAT GCA GCA ACA 3864
      E   Y   V   G   A   D   N   E   A   T   N   A   A   T

3865 TTC AAT GGA TAA GAA CAA TAA TTA AGC C    3892
      F   N   G   Z   E   Q   Z   L   S
```

FIG. 1A-6

```
 -15      MKLLL LNILLLCCLA DKLDEFSADN DYYDGGIMSR GKNAGSWYHS
  31 YTHQYDVFYY LAMQPWRHFV WTTCDKNDNT ECYKYTINED HNVKVEDINK
  81 TNIKQDFCQK EYAYPIEKYE VDWDNVPVDE QRIESVDING KTCFKYAAKR
 131 PLAYVYLNTK MTYATKTEAY DVCRMDFIGG RSITFRSFNT ENKAFIDQYN
 181 TNTTSKCLLN VYDNNVNTHL AIIFGITDST VIKSLQENLS LLSQLKTVKG
 231 VTLYYLKDDT YFTVNITLDQ LKYDTLVKYT AGTGQVDPLI NIAKNDLATK
 281 VADKSKDKNA NDKIKRGTMI VLMDTALGSE FNAETEFDRK NISVHTVVLN
 331 RNKDPKITRS ALRLVSLGPH YHEFTGNDEV NATITALFKG IRANLTERCD
 381 RDKCSGFCDA MNRCTCPMCC ENDCFYTSCD VETGSCIPWP KAKPKAKKEC
 431 PATCVGSYEC RDLEGCVVTK YNDTCQPKVK CMVPYCDNDK NLTEVCKQKA
 481 NCEADQKPSS DGYCWSYTCD QTTGFCKKDK RGKEMCTGKT NNCQEYVCDS
 531 EQRCSVRDKV CVKTSPYIEM SCYVAKCNLN TGMCENRLSC DTYSSCGGDS
 581 TGSVCKCDST TGNKCQCNKV KNGNYCNSKN HEICDYTGTT PQCKVSNCTE
 631 DLVRDGCLIK RCNETSKTTY WENVDCSNTK IEFAKDDKSE TMCKQYYSTT
 681 CLNGKCVVQA VGDVSNVGCG YCSMGTDNII TYHDDCNSRK SQCGNFNGKC
 731 IKGSDNSYSC VFEKDKTSSK SDNDICAECS SLTCPADTTY RTYTYDSKTG
 781 TCKATVQPTP ACSVCESGKF VEKCKDQKLE RKVTLENGKE YKYTIPKDCV
 831 NEQCIPRTYI DCLGNDDNFK SIYNFYLPCQ AYVTATYHYS SLFNLTSYKL
 881 HLPQSEEFMK EADKEAYCTY EITTRECKTC SLIETREKVQ EVDLCAEETK
 931 NGGVPFKCKN NNCIIDPNFD CQPIECKIQE IVITEKDGIK TTTCKNTTKA
 981 TCDTNNKRIE DARKAFIEGK EGIEQVECAS TVCQNDNSCP IITDVEKCNQ
1031 NTEVDYGCKA MTGECDGTTY LCKFVQLTDD PSLDSEHFRT KSGVELNNAC
1081 LKYKCVESKG SDGKITHKWE IDTERSNANP KPRNPCETAT CNQTTGETIY
1131 TKKTCTVSEF PTITPNQGRC FYCQCSYLDG SSVLTMYGET DKEYYDLDAC
1181 GNCRVWNQTD RTQQLNNHTE CILAGEINNV GAIAAATTVA AVIVAVVVAL
1231 IVVSIGLFKT YQLVSSAMKN AITITNENAE YVGADNEATN AATFNG
```

FIG. 1B

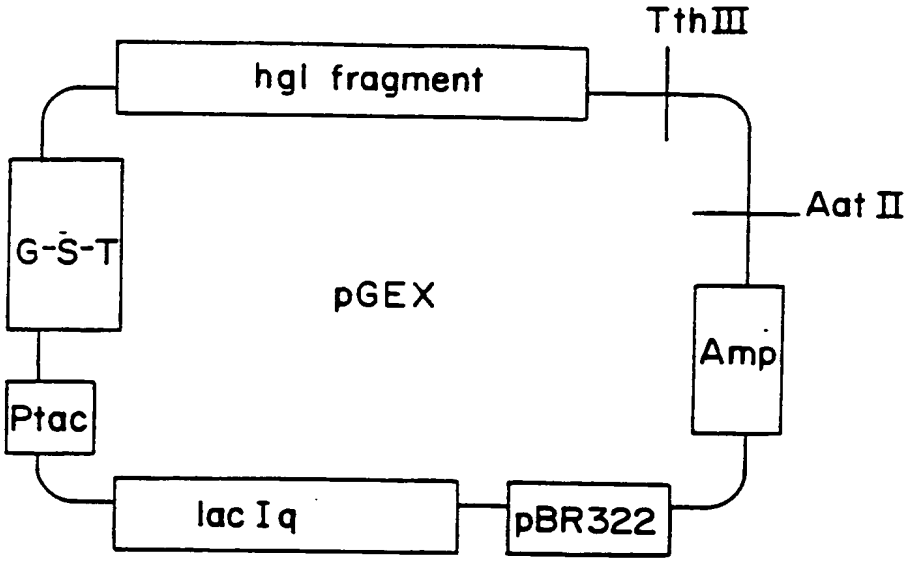

TthIII
hgl fragment
Aat II
G-S-T
pGEX
Amp
Ptac
lac Iq
pBR322

FIG. 2A

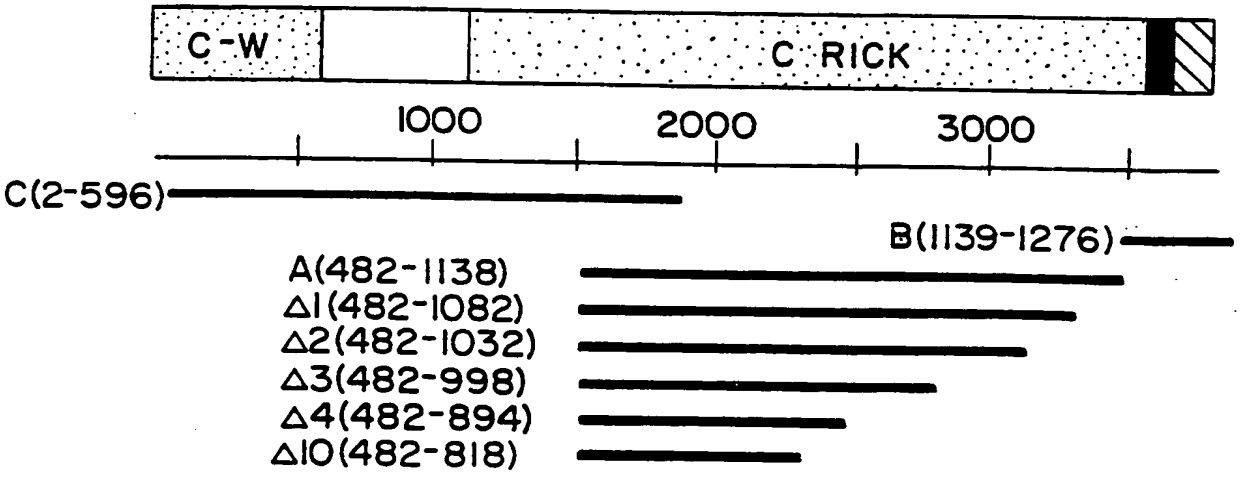

C-W
C RICK
1000
2000
3000
C(2-596)
B(1139-1276)
A(482-1138)
Δ1(482-1082)
Δ2(482-1032)
Δ3(482-998)
Δ4(482-894)
Δ10(482-818)

FIG. 2B

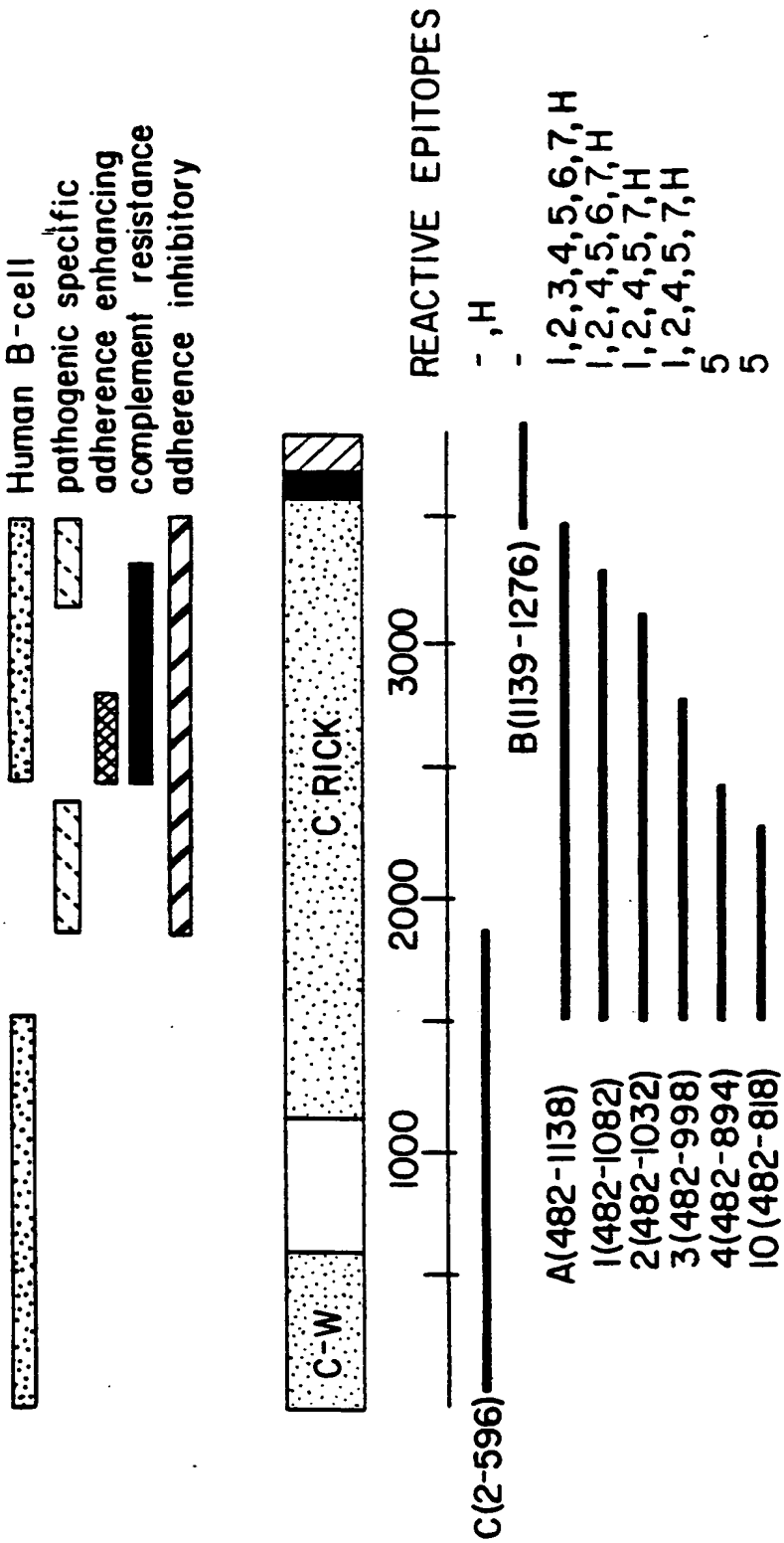

Human B-cell
pathogenic specific
adherence enhancing
complement resistance
adherence inhibitory
REACTIVE EPITOPES
C-W
C RICK
1000
2000
3000
C(2-596)
B(1139-1276)
-, H
-
1,2,3,4,5,6,7,H
1,2,4,5,6,7,H
1,2,4,5,7,H
1,2,4,5,7,H
5
5
A(482-1138)
1(482-1082)
2(482-1032)
3(482-998)
4(482-894)
10(482-818)

FIG. 3

```
  1 TTC TGT TAA ATA GGA AAG GCA AGT GAT TTA AAC AAG ACA ATG   42

 43 AAC TAG AAA GAC AAA GAT ATG AAA TTA TTA TTA TTA AAT ATC   84
                           M   K   L   L   L   L   N   I

 85 TTA TTA TTA TGT TGT CTT GCA GAT AAA CTT AAT GAA TTT TCA  126
      L   L   L   C   C   L   A   D   K   L   N   E   F   S

127 GCA GAT ATT GAT TAT TAT GAC CTT GGT ATT ATG TCT CGT GGA  168
      A   D   I   D   Y   Y   D   L   G   I   M   S   R   G

169 AAG AAT GCA GGT TCA TGG TAT CAT TCT TAT GAA CAT CAA TAT  210
      K   N   A   G   S   W   Y   H   S   Y   E   H   Q   Y

211 GAT GTT TTC TAT TAT TTA GCT ATG CAA CCA TGG AGA CAT TTT  252
      D   V   F   Y   Y   L   A   M   Q   P   W   R   H   F

253 GTA TGG ACT ACT TGT ACA ACA ACT GAT GGC AAT AAA GAA TGT  294
      V   W   T   T   C   T   T   T   D   G   N   K   E   C

295 TAT AAA TAT ACT ATC AAT GAA GAT CAT AAT GTA AAG GTT GAA  336
      Y   K   Y   T   I   N   E   D   H   N   V   K   V   E

337 GAT ATT AAT AAA ACA GAT ATT AAA CAA GAT TTT TGT CAA AAA  378
      D   I   N   K   T   D   I   K   Q   D   F   C   Q   K

379 GAA TAT GCA TAT CCA ATT GAA AAA TAT GAA GTT GAT TGG GAC  420
      E   Y   A   Y   P   I   E   K   Y   E   V   D   W   D

421 AAT GTT CCA GTT GAT GAA CAA CGA ATT GAA AGT GTA GAT ATT  462
      N   V   P   V   D   E   Q   R   I   E   S   V   D   I

463 AAT GGA AAA ACT TGT TTT AAA TAT GCA GCT AAA AGA CCA TTG  504
      N   G   K   T   C   F   K   Y   A   A   K   R   P   L

505 GCT TAT GTT TAT TTA AAT ACA AAA ATG ACA TAT GCA ACA AAA  546
      A   Y   V   Y   L   N   T   K   M   T   Y   A   T   K

547 ACT GAA GCA TAT GAT GTT TGT AGA ATG GAT TTC ATT GGA GGA  588
      T   E   A   Y   D   V   C   R   M   D   F   I   G   G

589 AGA TCA ATT ACA TTC AGA TCA TTT AAC ACA GAG AAT AAA GCA  630
      R   S   I   T   F   R   S   F   N   T   E   N   K   A

631 TTT ATT GAT CAA TAT AAT ACA AAC ACT ACA TCA AAA TGT CTT  672
      F   I   D   Q   Y   N   T   N   T   T   S   K   C   L

673 CTT AAA GTA TAT GAT AAT AAT GTT AAT ACA CAT CTT GCA ATT  714
      L   K   V   Y   D   N   N   V   N   T   H   L   A   I

715 ATC TTT GGT ATT ACT GAT TCT ACA GTC ATT AAA TCA CTT CAA  756
      I   F   G   I   T   D   S   T   V   I   K   S   L   Q
```

FIG. 4A

```
757 GAG AAC TTA TCT CTT TTA AAT AAA TTA ACA ACA GTC AAA GGA 798
    E   N   L   S   L   L   N   K   L   T   T   V   K   G

799 GTA ACA CTC TAC TAT CTT AAA GAT GAT ACT TAT TTT ACA GTT 840
    V   T   L   Y   Y   L   K   D   D   T   Y   F   T   V

841 AAT ATT ACT TTA AAT GAT TTG AAA TAT GAG ACA CTT GTC CAA 882
    N   I   T   L   N   D   L   K   Y   E   T   L   V   Q

883 TAC ACA GCA GGA ACA GGA CAA GTT GAT CCA CTT ATT AAT ATT 924
    Y   T   A   G   T   G   Q   V   D   P   L   I   N   I

925 GCT AAG AAT GAC TTA ACT GCT AAA GTT GCA GAT AAA AGT AAA 966
    A   K   N   D   L   T   A   K   V   A   D   K   S   K

967 GAT AAA AAT GCA AAT GAT AAA ATC AAA AGA GGA ACT ATG ATT 1008
    D   K   N   A   N   D   K   I   K   R   G   T   M   I

1009 GTG TTA ATG GAT ACT GCA CTT GGA TCA GAA TTT AAT GCG GAA 1050
     V   L   M   D   T   A   L   G   S   E   F   N   A   E

1051 ACA GAA TTT GAT AGA AAG AAT ATT TCA GTT CAT ACT GTT GTT 1092
     T   E   F   D   R   K   N   I   S   V   H   T   V   V

1093 CTT AAT AGA AAT AAA GAC CCA AAG ATT ACA CGT AGT GCA TTG 1134
     L   N   R   N   K   D   P   K   I   T   R   S   A   L

1135 AGA CTT GTT TCA CTT GGA CCA CAT TAT CAT GAA TTT ACA GGT 1176
     R   L   V   S   L   G   P   H   Y   H   E   F   T   G

1177 AAT GAT GAA GTT AAT GCA ACA ATC ACT GCA CTT TTC AAA GGA 1218
     N   D   E   V   N   A   T   I   T   A   L   F   K   G

1219 ATT AGA GCC AAT TTA ACA GAA AGA TGT GAT AGA GAT AAA TGT 1260
     I   R   A   N   L   T   E   R   C   D   R   D   K   C

1261 TCA GGA TTT TGT GAT GCA ATG AAT AGA TGC ACA TGT CCA ATG 1302
     S   G   F   C   D   A   M   N   R   C   T   C   P   M

1303 TGT TGT GAG AAT GAT TGT TTC TAT ACA TCC TGT GAT GTA GAA 1344
     C   C   E   N   D   C   F   Y   T   S   C   D   V   E

1345 ACA GGA TCA TGT ATT CCA TGG CCT AAA GCT AAA CCA AAA GCA 1386
     T   G   S   C   I   P   W   P   K   A   K   P   K   A

1387 AAG AAA GAA TGT CCA GCA ACA TGT GTA GGC TCA TAT GAA TGT 1428
     K   K   E   C   P   A   T   C   V   G   S   Y   E   C

1429 AGA GAT CTT GAA GGA TGT GTT GTT AAA CAA TAT AAT ACA TCT 1470
     R   D   L   E   G   C   V   V   K   Q   Y   N   T   S

1471 TGT GAA CCA AAA GTG AAA TGC ATG GTA CCA TAT TGT GAT AAT 1512
     C   E   P   K   V   K   C   M   V   P   Y   C   D   N
```

FIG. 4B

```
1513 GAT AAG AAT CTA ACT GAA GTA TGT AAA CAA AAA GCT AAT TGT 1554
     D   K   N   L   T   E   V   C   K   Q   K   A   N   C

1555 GAA GCA GAT CAA AAA CCA AGT TCT GAT GGA TAT TGT TGG AGT 1596
     E   A   D   Q   K   P   S   S   D   G   Y   C   W   S

1597 TAT ACA TGT GAC CAA ACT ACT GGT TTT TGT AAG AAA GAT AAA 1638
     Y   T   C   D   Q   T   T   G   F   C   K   K   D   K

1639 CGT GGT GAA AAT ATG TGT ACA GGA AAG ACA AAT AAC TGT CAA 1680
     R   G   E   N   M   C   T   G   K   T   N   N   C   Q

1681 GAA TAT GTT TGT GAT GAA AAA CAA AGA TGT ACT GTT CAA GAA 1722
     E   Y   V   C   D   E   K   Q   R   C   T   V   Q   E

1723 AAG GTA TGT GTA AAA ACA TCA CCT TAT ATT GAA ATG TCA TGT 1764
     K   V   C   V   K   T   S   P   Y   I   E   M   S   C

1765 TAT GTA GCC AAG TGT AAT CTC AAT ACA GGT ATG TGT GAG AAC 1806
     Y   V   A   K   C   N   L   N   T   G   M   C   E   N

1807 AGA TTA TCA TGT GAT ACA TAC TCA TCA TGT GGT GGA GAT TCT 1848
     R   L   S   C   D   T   Y   S   S   C   G   G   D   S

1849 ACA GGA TCA GTA TGT AAA TGT GAT TCT ACA ACT AAT AAC CAA 1890
     T   G   S   V   C   K   C   D   S   T   T   N   N   Q

1891 TGT CAA TGT ACT CAA GTA AAA AAC GGT AAT TAT TGT GAT TCT 1932
     C   Q   C   T   Q   V   K   N   G   N   Y   C   D   S

1933 AAT AAA CAT CAA ATT TGT GAT TAT ACA GGA AAA ACA CCA CAA 1974
     N   K   H   Q   I   C   D   Y   T   G   K   T   P   Q

1975 TGT AAA GTG TCT AAT TGT ACA GAA GAT CTT GTT AGA GAT GGA 2016
     C   K   V   S   N   C   T   E   D   L   V   R   D   G

2017 TGT CTT ATT AAG AGA TGT AAT GAA ACA AGT AAA ACA ACA TAT 2058
     C   L   I   K   R   C   N   E   T   S   K   T   T   Y

2059 TGG GAG AAT GTT GAT TGT TCT AAA ACT GAA GTT AAA TTC GCT 2100
     W   E   N   V   D   C   S   K   T   E   V   K   F   A

2101 CAA GAT GGT AAA TCT GAA AAT ATG TGT AAA CAA TAT TAT TCA 2142
     Q   D   G   K   S   E   N   M   C   K   Q   Y   Y   S

2143 ACT ACA TGT TTG AAT GGA CAA TGT GTT GTT CAA GCA GTT GGT 2184
     T   T   C   L   N   G   Q   C   V   V   Q   A   V   G

2185 GAT GTT TCT AAT GTA GGA TGT GGA TAT TGT TCA ATG GGA ACA 2226
     D   V   S   N   V   G   C   G   Y   C   S   M   G   T

2227 GAT AAT ATT ATT ACA TAT CAT GAT GAT TGT AAT TCA CGT AAA 2268
     D   N   I   I   T   Y   H   D   D   C   N   S   R   K
```

FIG. 4C

```
2269 TCA CAA TGT GGA AAC TTT AAT GGT AAG TGT GTA GAA AAT AGT 2310
      S   Q   C   G   N   F   N   G   K   C   V   E   N   S

2311 GAC AAA TCA TAT TCT TGT GTA TTT AAT AAG GAT GTT TCT TCT 2352
      D   K   S   Y   S   C   V   F   N   K   D   V   S   S

2353 ACA TCA GAT AAT GAT ATT TGT GCA AAA TGT TCT AGT TTA ACA 2394
      T   S   D   N   D   I   C   A   K   C   S   S   L   T

2395 TGT CCA GCT GAT ACT ACA TAC AGA ACA TAT ACA TAT GAC TCA 2436
      C   P   A   D   T   T   Y   R   T   Y   T   Y   D   S

2437 AAA ACA GGA ACA TGT AAA GCA ACT GTT CAA CCA ACA CCA GCA 2478
      K   T   G   T   C   K   A   T   V   Q   P   T   P   A

2479 TGT TCA GTA TGT GAA AGT GGT AAA TTT GTA GAA AAA TGC AAA 2520
      C   S   V   C   E   S   G   K   F   V   E   K   C   K

2521 GAT CAA AAA TTA GAA CGT AAA GTT ACT TTA GAA AAT GGA AAA 2562
      D   Q   K   L   E   R   K   V   T   L   E   N   G   K

2563 GAA TAT AAA TAC ACC ATT CCA AAA GAT TGT GTC AAT GAA CAA 2604
      E   Y   K   Y   T   I   P   K   D   C   V   N   E   Q

2605 TGC ATT CCA AGA ACA TAC ATA GAT TGT TTA GGT AAT GAT GAT 2646
      C   I   P   R   T   Y   I   D   C   L   G   N   D   D

2647 AAC TTT AAA TCT ATT TAT AAC TTC TAT TTA CCA TGT CAA GCA 2688
      N   F   K   S   I   Y   N   F   Y   L   P   C   Q   A

2689 TAT GTT ACA GCT ACC TAT CAT TAC AGT TCA TTA TTC AAT TTA 2730
      Y   V   T   A   T   Y   H   Y   S   S   L   F   N   L

2731 ACT AGT TAT AAA CTT CAT TTA CCA CAA AGT GAA GAA TTT ATG 2772
      T   S   Y   K   L   H   L   P   Q   S   E   E   F   M

2773 AAA GAG GCA GAC AAA GAA GCA TAT TGT ACA TAC GAA ATA ACA 2814
      K   E   A   D   K   E   A   Y   C   T   Y   E   I   T

2815 ACA AGA GAA TGT AAA ACA TGT TCA TTA ATT GAA ACT AGA GAA 2856
      T   R   E   C   K   T   C   S   L   I   E   T   R   E

2857 AAA GTC CAA GAA GTT GAT TTG TGT GCA GAA GAG ACT AAG AAT 2898
      K   V   Q   E   V   D   L   C   A   E   E   T   K   N

2899 GGA GGA GTT CCA TTC AAA TGT AAG AAT AAC AAT TGC ATT ATT 2940
      G   G   V   P   F   K   C   K   N   N   N   C   I   I

2941 GAT CCT AAC TTT GAT TGT CAA CCT ATT GAA TGT AAG ATT CAA 2982
      D   P   N   F   D   C   Q   P   I   E   C   K   I   Q

2983 GAG ATT GTT ATT ACA GAA AAA GAT GGA ATA AAA ACA ACA ACA 3024
      E   I   V   I   T   E   K   D   G   I   K   T   T   T
```

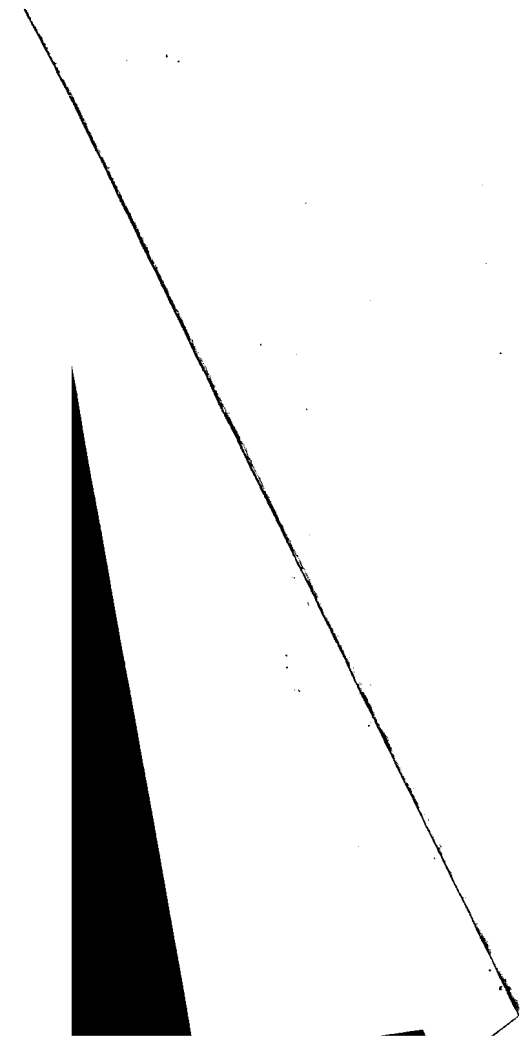
FIG. 4D

```
3025 TGT AAA AAT ACC ACA AAA ACA ACA TGT GAC ACT AAC AAT AAG 3066
       C   K   N   T   T   K   T   T   C   D   T   N   N   K

3067 AGA ATA GAA GAT GCA CGT AAA GCA TTC ATT GAA GGA AAA GAA 3108
       R   I   E   D   A   R   K   A   F   I   E   G   K   E

3109 GGA ATT GAG CAA GTA GAA TGT GCA AGT ACT GTT TGT CAA AAT 3150
       G   I   E   Q   V   E   C   A   S   T   V   C   Q   N

3151 GAT AAT AGT TGT CCA ATT ATT ACT GAT GTA GAA AAA TGT AAT 3192
       D   N   S   C   P   I   I   T   D   V   E   K   C   N

3193 CAA AAC ACA GAA GTA GAT TAT GGA TGT AAA GCA ATG ACA GGA 3234
       Q   N   T   E   V   D   Y   G   C   K   A   M   T   G

3235 GAA TGT GAT GGT ACT ACA TAT CTT TGT AAA TTT GTA CAA CTT 3276
       E   C   D   G   T   T   Y   L   C   K   F   V   Q   L

3277 ACT GAT GAT CCA TCA TTA GAT AGT GAA CAT TTT AGA ACT AAA 3318
       T   D   D   P   S   L   D   S   E   H   F   R   T   K

3319 TCA GGA GTT GAA CTT AAC AAT GCA TGT TTG AAA TAT AAA TGT 3360
       S   G   V   E   L   N   N   A   C   L   K   Y   K   C

3361 GTT GAG AGT AAA GGA AGT GAT GGA AAA ATC ACA CAT AAA TGG 3402
       V   E   S   K   G   S   D   G   K   I   T   H   K   W

3403 GAA ATT GAT ACA GAA CGA TCA AAT GCT AAT CCA AAA CCA AGA 3444
       E   I   D   T   E   R   S   N   A   N   P   K   P   R

3445 AAT CCA TGC GAA ACC GCA ACA TGT AAT CAA ACA ACT GGA GAA 3486
       N   P   C   E   T   A   T   C   N   Q   T   T   G   E

3487 ACT ATT TAC ACA AAG AAA ACA TGT ACT GTT TCA GAA GAA TTC 3528
       T   I   Y   T   K   K   T   C   T   V   S   E   E   F

3529 CCA ACA ATC ACA CCA AAT CAA GGA AGA TGT TTC TAT TGT CAA 3570
       P   T   I   T   P   N   Q   G   R   C   F   Y   C   Q

3571 TGT TCA TAT CTT GAC GGT TCA TCA GTT CTT ACT ATG TAT GGA 3612
       C   S   Y   L   D   G   S   S   V   L   T   M   Y   G

3613 GAA ACA GAT AAA GAA TAT TAT GAT CTT GAT GCA TGT GGT AAT 3654
       E   T   D   K   E   Y   Y   D   L   D   A   C   G   N

3655 TGT CGT GTT TGG AAT CAG ACA GAT AGA ACA CAA CAA CTT AAT 3696
       C   R   V   W   N   Q   T   D   R   T   Q   Q   L   N

3697 AAT CAC ACC GAG TGT ATT CTC GCA GGA GAA ATT AAT AAT GTT 3738
       N   H   T   E   C   I   L   A   G   E   I   N   N   V

3739 GGA GCT ATT GCA GCG GCA ACT ACT GTG GCT GTA GTT GTA GTT 3780
       G   A   I   A   A   A   T   T   V   A   V   V   V   V
```

FIG. 4E

```
3781 GCA GTC GTA GTT GCA TTA ATT GTT GTT TCT ATT GGA TTA TTT 3822
       A   V   V   V   A   L   I   V   V   S   I   G   L   F

3823 AAG ACT TAT CAA CTT GTT TCA TCA GCT ATG AAG AAT GCC ATT 3864
       K   T   Y   Q   L   V   S   S   A   M   K   N   A   I

3865 ACA ATA ACT AAT GAA AAT GCA GAA TAT GTT GGA GCA GAT AAT 3906
       T   I   T   N   E   N   A   E   Y   V   G   A   D   N

3907 GAA GCA ACT AAT GCA GCA ACA TTC AAT GGA TAA GAA CAA TAA 3948
       E   A   T   N   A   A   T   F   N   G   Z

3949 TTA AGA GAA TTG AAT AAC ATT TTA TGT TTT TAG ATT AAA AAT 3990

3991 AAA AAG AAG AAT AAA TTG AGT GAT AAA CAA TGA ATA AAA TAA 4032

4033 ATA AAA ATA AAC AAG AAT AAA GTG AAC ATC ATT TTT ATT TTC 407

4075 ATA TTT TAA CAA CAC T 4090
```

FIG. 4F



```
-15 MKLLLLNILL LCCLADKLNE FSADIDYYDL GIMSRGKNAG SWYHSYEHQY DVFYYLAMQP WRHFVWTTCT TTDGNKECYK
 66 YTINEDHNVK VEDINKTDIK QDFCQKEYAY PIEKYEVDWD NVPVDEQRIE SVDINGKTCF KYAAKRPLAY VYLNTKMTYA
146 TKTEAYDVCR MDFIGGRSIT FRSFNTENKA FIDQYNTNTT SKCLLKVYDN NVNTHLAIIF GITDSTVIKS LQENLSLLNK
226 LTTVKGVTLY YLKDDTYFTV NITLNDLKYE TLVQYTAGTG QVDPLINIAK NDLTAKVADK SKDKNANDKI KRGTMIVLMD
306 TALGSEFNAE TEFDRKNISV HTVVLNRNKD PKITRSALRL VSLGPHYHEF TGNDEVNATI TALFKGIRAN LTERCDRDKC
386 SGFCDAMNRC TCPMCCENDC FYTSCDVETG SCIPWPKAKP KAKKECPATC VGSYECRDLE GCVVKQYNTS CEPKVKCMVP
466 YCDNDKNLTE VCKQKANCEA DQKPSSDGYC WSYTCDQTTG FCKKDKRGEN MCTGKTNNCQ EYVCDEKQRC TVQEKVCVKT
546 SPYIEMSCYV AKCNLNTGMC ENRLSCDTYS SCGGDSTGSV CKCDSTTNNQ CQCTQVKNGN YCDSNKHQIC DYTGKTPQCK
626 VSNCTEDLVR DGCLIKRCNE TSKTTYWENV DCSKTEVKFA QDGKSENMCK QYYSTTCLNG QCVVQAVGDV SNVGCGYCSM
706 GTDNIITYHD DCNSRKSQCG NFNGKCVENS DKSYSCVFNK DVSSTSDNDI CAKCSSLTCP ADTTYRTYTY DSKTGTCKAT
786 VQPTPACSVC ESGKFVEKCK DQKLERKVTL ENGKEYKYTI PKDCVNEQCI PRTYIDCLGN DDNFKSIYNF YLPCQAYVTA
866 TYHYSSLFNL TSYKLHLPQS EEFMKEADKE AYCTYEITTR ECKTCSLIET REKVQEVDLC AEETKNGGVP FKCKNNNCII
946 DPNFDCQPIE CKIQEIVITE KDGIKTTTCK NTTKTTCDTN NKRIEDARKA FIEGKEGIEQ VECASTVCQN DNSCPIITDV
1026 EKCNQNTEVD YGCKAMTGEC DGTTYLCKFV QLTDDPSLDS EHFRTKSGVE LNNACLKYKC VESKGSDGKI THKWEIDTER
1106 SNANPKPRNP CETATCNQTT GETIYTKKTC TVSEEFPTIT PNQGRCFYCQ CSYLDGSSVL TMYGETDKEY YDLDACGNCR
1186 VWNQTDRTQQ LNNHTECILA GEINNVGAIA AATTVAVVVV AVVVALIVVS IGLFKTYQLV SSAMKNAITI TNENAEYVGA
1266 DNEATNAATF NG
```

FIG. 5

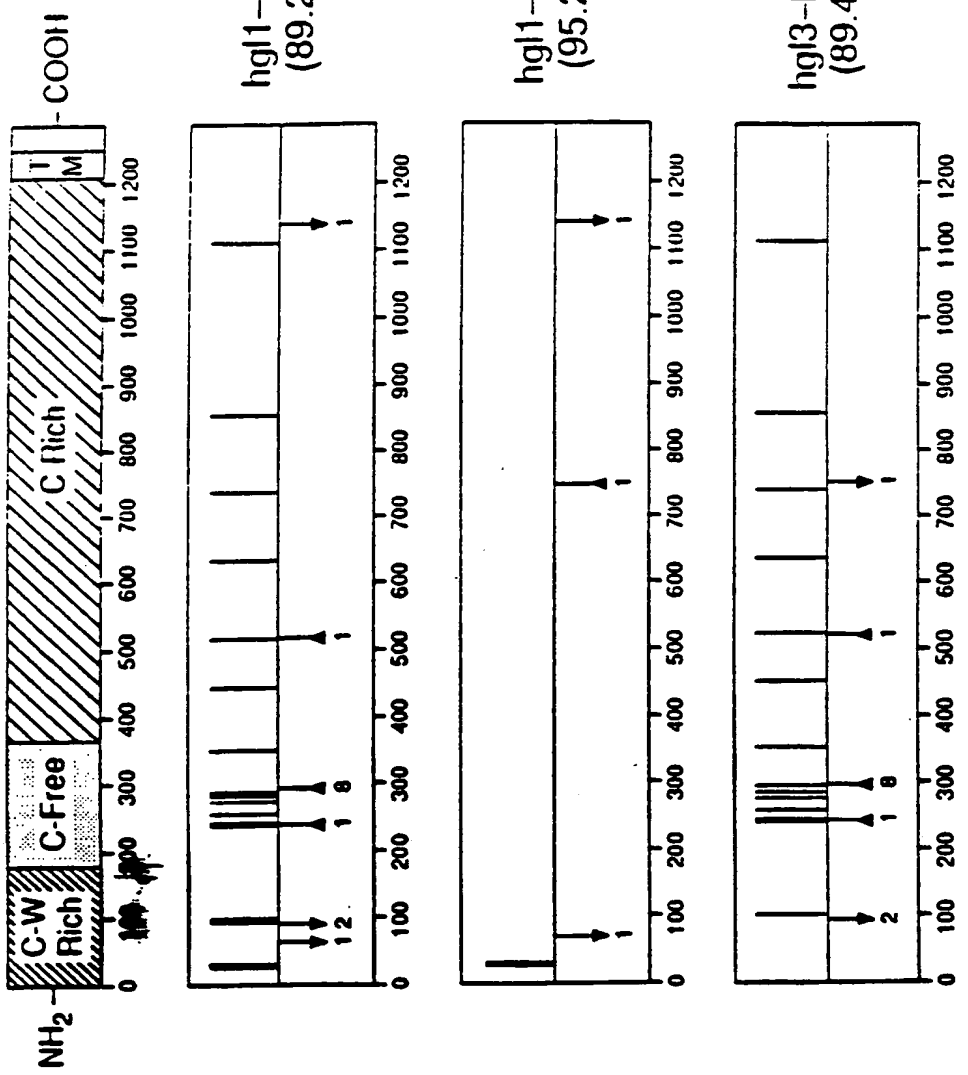

NH2
C-W Rich
C-Free
C Rich
T
M
COOH
hgl1–hgl2 (89.2%)
hgl1–hgl3 (95.2%)
hgl3–hgl2 (89.4%)
0 100 200 300 400 500 600 700 800 900 1000 1100 1200

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/06890

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :G01N 33/569; C07K 13/00; C12N 15/30
US CL :435/7.22; 530/396; 424/88
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.22, 7.92, 7.93, 970, 975; 530/395, 396; 424/88; 436/518, 514/8; 536/23.4, 23.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, EMBASE, BIOSIS, CA SEARCH, WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X – Y	J. Clin. Microbiol., Volume 30, No. 11, issued November 1992, Zhang et al., "Use of a Recombinant 170-Kilodalton Surface Antigen of *Entamoeba histolytica* for Serodiagnosis of Amebiasis and Identification of Immunodominant Domains of the Native Molecule", pages 2780-2792, especially see page 2788, col. 2, lines 4-16 and page 2791, first paragraph.	1-3, 9, 13-14, 16 4-8, 10-12, 15, 17-29
Y	Proc. Natl. Acad. Sci. USA, Volume 88, issued March 1991, Tannich et al., "Primary structure of the 170-kDa surface lectin of pathogenic *Entamoeba histolytica*", pages 1849-1853, especially Fig. 3 on page 1851.	1-29

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
12 SEPTEMBER 1994

Date of mailing of the international search report
20 SEP 1994

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
CAROL E. BIDWELL
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/06890

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO, A, 91/12529 (PETRI ET AL.) 22 August 1991, see entire document, especially see Abstract, page 18, lines 5-12,	1-29
Y	PLOTKIN ET AL., "VACCINES", published 1988 by W.B. Saunders Company (Philadelphia), pages 568-575, especially see page 572, first full paragraph to page 573, line 4 and Table 29-6.	18-28
Y	US, A, 5,004,608 (RAVDIN ET AL.) 02 April 1991, see entire document, especially claims 1 and 5.	18-28
Y	Proc. Natl. Acad. Sci. USA, Volume 88, issued April 1991, Mann et al., "Sequence of a cysteine-rich galactose-specific lectin of *Entamoeba histolytica*", pages 3248-3252, especially see Abstract on page 3248 and Fig. 1.	1-29
Y	JAMA, Volume 266, No. 14, issued 09 October 1991, Stanley et al., "Serodiagnosis of Invasive Amebiasis Using a Recombinant *Entamoeba histolytica* Protein", pages 1984-1986, especially see Abstract on page 1984.	1-29
X, P	Mol. Biochem. Parasitol., Volume 62, No. 1, issued November 1993, Purdy et al., "Analysis of the gene family encoding the *Entamoeba histolytica* galactose-specific adhesin 170-kDa subunit", pages 53-60, especially see Abstract.	29